

Pan-tissue Transcriptome Analysis Reveals Sex-dimorphic Human Aging

Reviewed Preprint

v2 • July 9, 2025

Revised by authors

Reviewed Preprint

v1 • December 3, 2024

Siqi Wang , Danyue Dong, Xin Li, Zefeng Wang 

Shanghai Institute of Nutrition and Health, University of Chinese Academy of Sciences, Chinese Academy of Sciences, Shanghai, China • Department of Integrative Biology and Physiology, University of California, Los Angeles, Los Angeles, United States • School of Life Science, Southern University of Science and Technology, Shenzhen, China

 https://en.wikipedia.org/wiki/Open_access

 Copyright information

eLife Assessment

In this study Wang et. al. mined publicly available RNA-seq data from The Genotype-Tissue Expression (GTEx) database spanning multiple tissues to ask the question of how transcriptomes are changed with age and in both sexes. The authors provide **solid** evidence reporting widespread gene expression changes and alternative splicing events that vary in an age- and sex-dependent manner. An **important** finding is that many of these changes coincide with the time sex hormones begin to decline; additionally, the rate of aging is faster in males than in females.

<https://doi.org/10.7554/eLife.102449.2.sa3>

Abstract

Complex diseases often exhibit sex-dimorphism in morbidity and prognosis, many of which are age-related. However, the underlying mechanisms of sex-dimorphic aging remain foggy, with limited studies across multiple tissues. We systematically analyzed ~17,000 transcriptomes from 35 human tissues to quantitatively evaluate the individual and combined contributions of sex and age to transcriptomic variations. We discovered extensive sex-dimorphisms during aging with distinct patterns of change in gene expression and alternative splicing (AS). Intriguingly, the male-biased age-associated AS events have a stronger association with Alzheimer's disease, and the female-biased events are often regulated by several sex-biased splicing factors that may be controlled by estrogen receptors. Breakpoint analysis showed that sex-dimorphic aging rates are significantly associated with decline of sex hormones, with males having a larger and earlier transcriptome change. Collectively, this study uncovered an essential role of sex during aging at the molecular and multi-tissue levels, providing insight into sex-dimorphic regulatory patterns.

Introduction

Human diseases often exhibit differences between females and males, including sex-differential morbidity, prognosis, and mortality ^{1–3}. The sex-dimorphism has been widely reported in neurological disorders ^{4,5}, cardiovascular diseases ⁶, immunological defects ⁷, and cancers ⁸. Accordingly, the life expectancy shows substantial variability between females and males. The evidence on sex-differential mortality was recently reported during the COVID-19 pandemic, mainly affected by different immune responses between females and males ^{9,10}. The known molecular mechanisms mostly revolved around sex-differential genetic variants, epigenetics, transcriptomes, and sex-differential responses to environmental exposures ². It is worth noting that many diseases with sex-dimorphism are also age-related, especially in neurodegenerative disorders, cardiovascular diseases, and cancers ¹¹. Such sex-differential susceptibility contributes substantially to different life expectancies between females and males, and the related underlying mechanisms are thought to be hormone-driven, mitochondria-related, and sex-chromosome-linked ^{11–15}. Previously, sex-differential aging signatures, including chronological trends and gene networks, have been studied in the whole blood and brain regions of healthy donors ¹⁶, revealing the sex differences of disease vulnerabilities during aging ^{17,18}. Sex differences regarding the composition and inflammatory signaling of immunocytes in blood were also analyzed in single-cell resolution, which suggested sex-differential aging in the immune response ¹⁹. Moreover, at proteogenomic level, sex-biased genes play key roles in several important cellular processes during cardiac aging, such as mitochondrial metabolism, RNA splicing, and translation, implying sex dimorphism in cardiac diseases ²⁰. However, a systematic analysis across multiple tissues on the sex-differential aging and underlying molecular mechanisms is currently lacking.

The Genotype-Tissue Expression (GTEx) project contains a large set of high-throughput sequencing data from postmortem donors across 54 human tissues ^{21,22}. Previous studies have uncovered many sex-biased genes or expression quantitative trait loci (eQTLs) ^{22,23} across different tissues that could affect several critical cellular signal pathways related to human diseases. Since the GTEx donors have a wide age range of 20–70 years, it is possible to identify genes and molecular pathways significantly changed during aging process ^{24,25}. However, the relationship between sex-differential regulation and aging has not been thoroughly studied on a transcriptome or pan-tissue level scale. Therefore, the GTEx data could also provide an integrated resource for a comprehensive study of this topic.

In addition to transcriptional regulation, most human genes undergo alternative splicing (AS) to produce different isoforms with distinct activities ²⁶. AS is tightly regulated in various tissues or developmental stages by multiple *cis*-acting elements and *trans*-acting factors ^{27,28}. Dysregulation of AS is closely related to many age-related diseases, such as neurodegenerative diseases ²⁹ and cancers ³⁰. Since most introns are co-transcriptionally spliced, AS is also affected by various factors that regulate transcription, such as promoter activity, Pol II elongation, and chromatin modification/remodelling ²⁸. Recent studies reported that AS regulation is closely associated with sex ³¹ and age ^{32–34}; thus, it is intriguing to further explore the AS regulation during aging stages in different sexes.

In this study, we initiated an integrated analysis of pan-tissue transcriptome data in GTEx, and systematically determined how sex and age contributed to the global variations of gene expression (GE) and AS (**Fig. 1**). We further focused on the sex-biased AS changes during aging, and found that these AS events are significantly associated with neurodegenerative diseases in a sex-biased fashion. We further examined the chronological changes of age-associated genes, and found a sex-

dimorphic transcriptome aging, with males showing an earlier onset of aging and a faster aging rate in most tissues. Collectively, this study clarified the sex and age effects on the transcriptomic changes, revealing the sex-dimorphic aging at a multi-tissue level and their potential mechanisms.

Results

Sex and age are critical drivers to the global transcriptome variation

To systematically study how sex and age affect transcriptome complexity, we conducted a thorough analysis of RNA-seq data from GTEx project²¹ to quantify the GE and AS patterns of all genes from 54 tissues (**Fig. 1**). Due to data sparseness, the brain tissues were recombined into four functional regions (table S1), including hormone- or emotion-related region, movement-related region, memory-related region, and decision-related region (See Methods). We performed the principal component analysis (PCA) on GE and AS data, and observed the sex and age differences in several tissues (fig. S1). Based on a previous study³⁵, we designed a method called principal component-based signal-to-variation ratio (pcSVR) to measure the adjusted Euclidean distance in higher dimensions. The pcSVR quantified the distance between different sex or age groups divided by the data dispersion within each group (intrinsic differences between groups will result in a pcSVR value significantly larger than 1), serving as a reliable measurement for the sex or age effects on transcriptomic variations (fig. S2A, see Materials and Methods, Equations 1–3). Compared with the direct identification of differentially expressed genes or AS events, pcSVR provides a global measurement by considering variations from all genes and AS events between different groups.

We grouped the samples into male vs. female and young vs. old, and calculated the sex-pcSVR and age-pcSVR in multiple tissues (**Fig. 2**, A and B). Due to a large variation of menopause ages among different individuals and a continuous decline of sexual steroid levels between ages 40 and 60^{36,37}, we grouped the samples into young (age<40) vs. old (age>60) with an age gap instead of a specific age cutoff to reduce the data noise. We found that the breast tissue showed the largest sex-pcSVR, while the adipose and cardiovascular tissues showed larger age-pcSVR (i.e., differences between age groups). The tissues with significant sex differences were similar to those reported in a previous study of sex-differential GE using transcriptomic signal-to-noise ratio³⁵, including breast, thyroid, skin, and adipose. Using the permutation test for pcSVRs, we found that age showed substantially larger effects than sex to human transcriptome in most tissues as judged by both GE and AS (**Fig. 2**, A and B, table S2). Moreover, AS was significantly affected by both sex and age across most tissues, while GE was affected by sex in a much smaller number of tissues as compared to AS profiles (i.e., comparing the orange triangles in **Fig. 2A** vs. 2B). For example, the coronary artery and adrenal gland are significantly affected by sex and age in their AS profile, but their GE profiles are not affected by sex or age (**Fig. 2**, A and B). Similar results were found even when removing the genes and AS events encoded by sex-chromosomes (fig. S2B and C), selecting a range of PC cutoffs to capture different global variance (fig. S2, D and E), or using other approaches for p-value estimation (e.g., boot-strapping with replacement, see Materials and Methods, fig. S2, F and G).

We further compared the age and sex effects calculated by GE vs. AS (table S2), and found a significant correlation between the age-pcSVR of GE and AS in most tissues ($R=0.67$, $p<0.05$, **Fig. 2C**), suggesting the age contribution to transcriptome variation is largely consistent regardless of the input data types (i.e., GE or AS). However, the sex effects on GE and AS showed a much weaker correlation ($R=0.21$, **Fig. 2D**), consistent with our earlier results (**Fig. 2**, A and B). We also found that age was a more vital driver of AS variation than GE in specific tissues like coronary artery and stomach, whereas sex contributed more to AS variations in whole blood, skeletal muscle, and most brain regions.

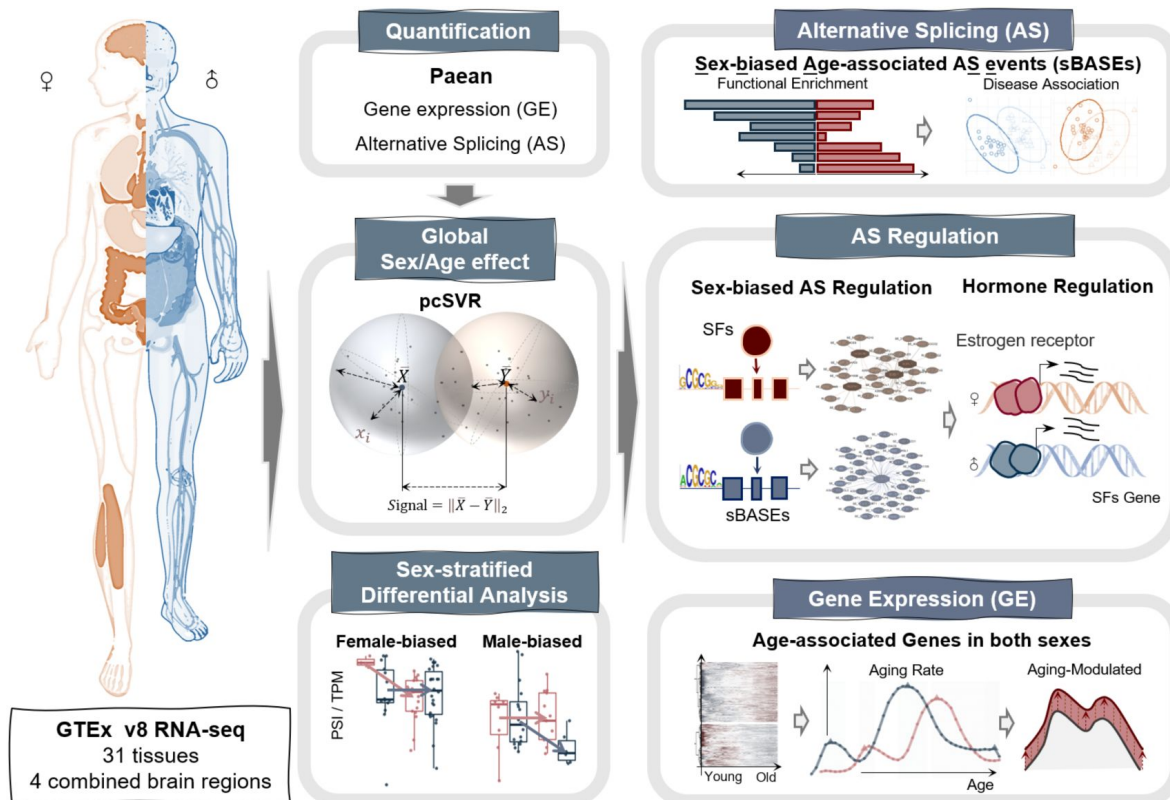


Fig. 1

Schematic Overview of our study.

RNA-seq datasets from 54 human tissues (including 13 brain regions) were downloaded from GTEx Portal. GE and AS quantifications were performed by Paeon. The individual sex or age and age-by-sex effects on global transcriptomic variation were evaluated by designing a pcSVR value, and the age-differential genes/AS events were subsequently identified in each sex. Sex-biased age-associated AS events were used to find the sex-specific associations with certain diseases. We also constructed AS regulatory networks to find the regulatory age-associated splicing factors (SF) and explored the transcriptional regulation of sex hormones via nuclear receptors. Moreover, we focused on age-associated GE and then evaluate the aging rate and their modulated genes in different sexes.

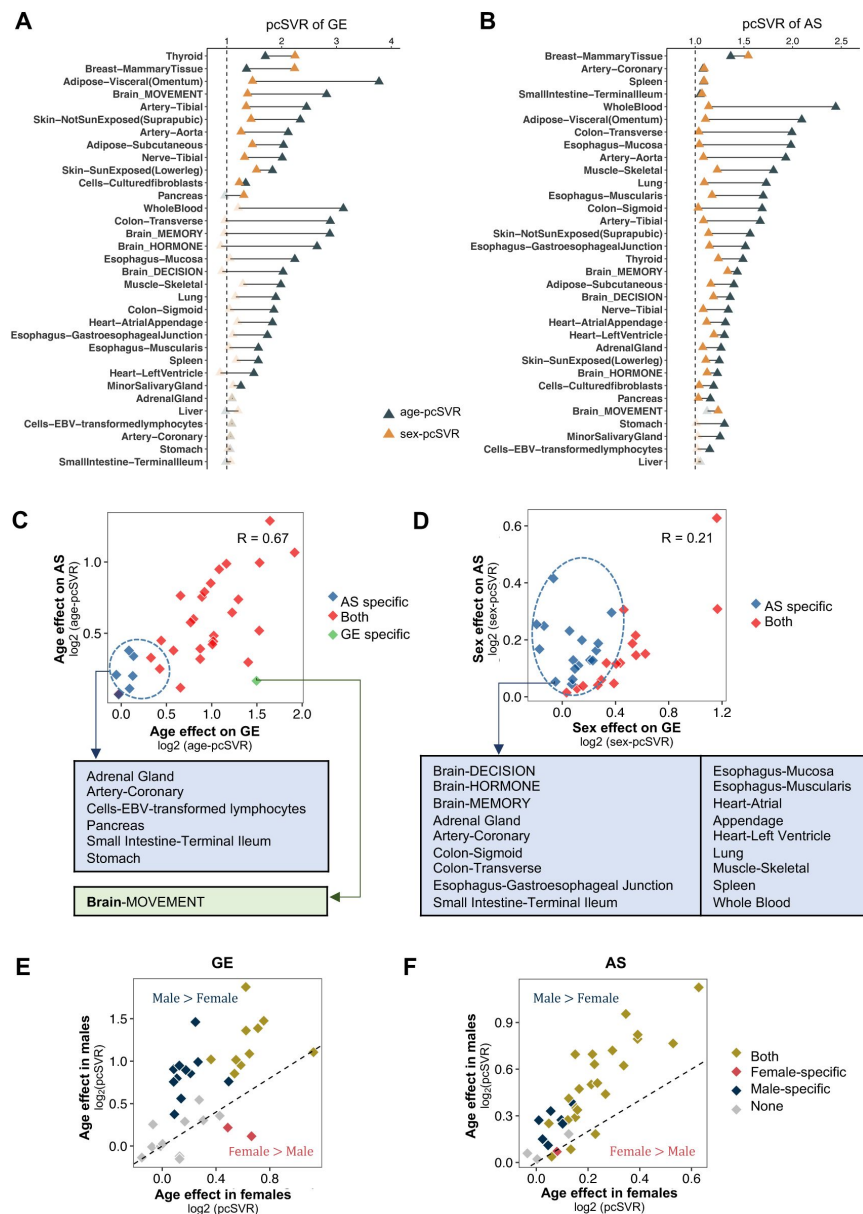


Fig. 2

Individual sex or age effects and combined age-by-sex effects on global transcriptome variation.

(A-B) pcSVR between different sex or age groups calculated by GE (a) and AS (b). The yellow triangles show the significant sex-pcSVR calculated between females and males, while black triangles show the significant age-pcSVR calculated between young (i.e., 20-40 years old) and old (i.e., >60 years old) using the empirical p-value cutoff 0.1. Insignificant data points are labeled in grey.

(C) The age effect on transcriptome variation as calculated by GE vs. AS. The X-axis of the scatter plot show the $\log_2$ transformed age-pcSVR between young and old calculated by GE, while the Y-axis shows the values calculated by AS. Blue dots indicate the tissues with significant pcSVR specific in AS, while green dots indicate the tissues with significant pcSVR specific in GE. Tissue names in each class are labeled under the scatter plot.

(D) The sex effect on transcriptome variation as calculated by GE vs. AS. The analyses and labels are similar to panel c except the sex-pcSVR is used in all calculations.

(E-F) Scatter plots of the age effect between females vs. males calculated by GE (E) and AS (F). The X-axis shows $\log_2$ transformed age-pcSVR calculated in females, while the Y-axis shows the age effect calculated in males. Dashed lines showed equal age-pcSVR in females and males. Dark red and blue dots indicate the tissues with female-specific and male-specific age effects.

To further examine the sex-differential age effect, we compared the age-pcSVR between females and males (**Fig. 2**, E and F; GE and AS, respectively). Consistent with earlier findings (**Fig. 2C**), we found significant age effects in both sexes across most tissues. Moreover, the age effects in both GE and AS were apparently larger in males than in females (**Fig. 2**, E and F). These results suggested that age and sex contributed significantly to transcriptome variation in various tissues with certain degrees of heterogeneity for different data types.

Differential GE and AS analyses reveal general interactions of age-by-sex

We next identified specific genes or AS events responsible for sex or age effect on transcriptomic variations via a linear regression model containing an age-by-sex coefficient term (see Materials and Methods, **Equations 4** and **5**). In each tissue, we identified the sex- or age-differential genes and AS events (fig. S3, A and B). Consistent with previous reports^{35,38,39}, the breast tissue contains the largest amount of sex-differentially expressed genes, while the age-biased genes were widely discovered in brain regions and artery-related tissues. Interestingly, in most tissues, there were no significant overlaps between age-differential genes and AS events (fig. S3A, hypergeometric test p-value < 0.01). In contrast, we found significant overlaps between sex-differential genes and AS events in most tissues (fig. S3B).

As a control, we constructed a separate regression model to evaluate the contribution of sampling biases, including sample sizes and the total number of genes or AS events detected in each tissue (**Equation 6**). We found that the sample size did not significantly affect the number of identified sex-differential or age-differential genes/AS events (fig. S3, C and D), suggesting our analyses are not affected by sampling biases. Furthermore, using the coefficient of the age-by-sex term in this model allowed us to identify thousands of genes and AS events affected by the functional interactions between sex and age (i.e., the sex effects on GE/AS depend on the age, *or vice versa*) (fig. S3E), suggesting extensive sex-dimorphic changes of the transcriptome during aging.

A recent study reported the environmental effect on human genetic background during aging⁴⁰, which should be considered when analyzing large datasets from different populations. In the linear regression model, we included the surrogate variables (SV) as confounders to control for the effect of genetic backgrounds (**Equations 4** and **5**). Using the differential GE analysis in whole blood as an example (other tissues have the same trend), we evaluated the correlation between SVs with donors' ethnicity using Point Biserial Correlation and found that many SVs showed significant correlations with the donors' ethnical background (fig. S4A), suggesting that these SVs could effectively reflect the variables of genetic background. Additionally, we calculated the correlation between SVs with the top 5 principal components as judged by whole genome sequencing (WGS) data (fig. S4B), and found roughly the same set of SVs also showed a significant correlation with PC1. Moreover, recent studies demonstrated the powerful impacts of post-mortem interval (PMI) and time of death (TOD), which included the death seasons and day times on gene expression^{41,42}. Thus, we carefully evaluated whether both factors are controlled as potential confounders in our linear model. Our results showed that PMI and TOD significantly correlate with covariates in most tissues (fig. S4C-E), suggesting that their effects could be effectively regressed in our model. Together, the SVs in the linear regression models could effectively control the cofounders of human genetic regulation.

To further evaluate the robustness of the differential analyses, we performed detailed assessments of the linear regression models and the selection of GE/AS cutoffs using the decision-related brain region as an example. We calculated the fraction of original differential genes/AS events identified in each permutation iteration, and the false discovery rate (FDR) for each gene/AS event using the permutation analysis in the linear regression model (see Materials and Methods). Briefly, we generated the permuted data 1,000 times by randomly shuffling the group labels (sex and age) with the same sample sizes as the original labels. The distribution of the p-values of sex/age-

differential genes/AS events identified based on the original and each shuffled phenotype data were evaluated (fig. S5, A and D). We first observed low fractions of sex/age-differential genes/AS events in most iterations under our cutoffs (fig. S5, B and E). We found that FDRs of most sex/age-differential genes/AS events were not more than 5% using the linear regression model (fig. S4, C and F), which is much lower than the FDR of DESeq2 and edgeR⁴³. As a result, this approach could justifiably control the false discovery for differential analysis. Furthermore, the TPMs of most sex- or age-differential genes were higher than 1, and the junction read counts (JC) of most differential AS events were higher than 10 (fig. S6, A and B), suggesting that our cutoffs (TPM > 1 or JC > 10) preserved most of the differential GE and AS events. In addition, we found a low overlap between the genes with age-differential GE and AS under most combinations of TPM and JC cutoffs in the decision-related brain region (fig. S6C), which is consistent with our earlier findings (fig. S3A). Collectively, these controls confirmed the reliability of our differential analyses.

AS regulation shows a stronger sex-dimorphism during aging than GE across all tissues

We further studied the transcriptomic changes during aging process in different sexes by identifying age-associated genes or AS events separately in females and males (table S3). We found that the age-associated genes significantly overlapped between females and males in most tissues (p -values < 0.01 by hypergeometric tests, **Fig. 3A**), whereas the overlaps of the age-associated AS events were much smaller (**Fig. 3B**). Consistently, in most tissues, the age effect on GE was more correlated between females and males (Spearman's correlation between β_F and β_M , **Equation 7**) than the age effect on AS (**Fig. 3C**), regardless of different GE/AS filters used in such analysis (fig. S6D). Taken together, these observations suggested that the age effect on GE is more consistent in both sexes, while age-associated AS showed a stronger discrepancy between two sexes, implying that splicing regulation had a stronger sex-dimorphism during aging.

AS changes show a male-biased association with Alzheimer's disease in human brain

Because most age-associated AS events were sex-specific, we defined the AS events significantly affected by age in only one sex as the sex-biased age-associated AS events (sBASEs), and further examined their functional impacts on human health using Gene Ontology and Disease Ontology analyses. The functional association of sBASEs varied considerably in different tissues and sexes (fig. S7, table S4). Having established patterns across all tissues, we next focused on tissue-specific AS changes in the brain that is highly affected by age-related processes. Since many neurological conditions show sex differences, the sex-biased splicing events in the brain could be helpful to explain differential susceptibility to age-related cognitive decline and neurodegeneration. Notably, the sBASEs in brain regions were significantly enriched in pathways such as amyloid-beta formation and cytoskeleton (**Fig. 4A**), with the male decision-related brain region being particularly associated with cognitive and psychotic disorders (fig. S8A). Because previous studies suggested that the disease manifestations of Alzheimer's Disease (AD) showed sex differences in cognitive decline and brain atrophy⁴⁴, we next focused on sBASEs in brain to explore their contribution to the sex-dimorphic risks of neurodegenerative diseases. This question was addressed by an integrative analysis of the large datasets from other independent AD studies, which is a common strategy (and strength) of computation biology in the big data era.

We analyzed the RNA-seq dataset of the brain prefrontal cortex with a dataset of 91 AD patients and healthy controls manually curated to match the sex and disease ratio (Female:Male≈1:1; AD:Control≈1:1, see Materials and Methods). Surprisingly, although the number of sBASEs in males was smaller compared with that in females, we found that the AD-related AS events significantly overlapped with the sBASEs in males ($p = 0.037$) but not in females ($p = 0.425$) (**Fig. 4B**). In addition, the magnitude of AD-associated AS changes across the male-specific sBASEs (i.e., PSI_{AD}^-

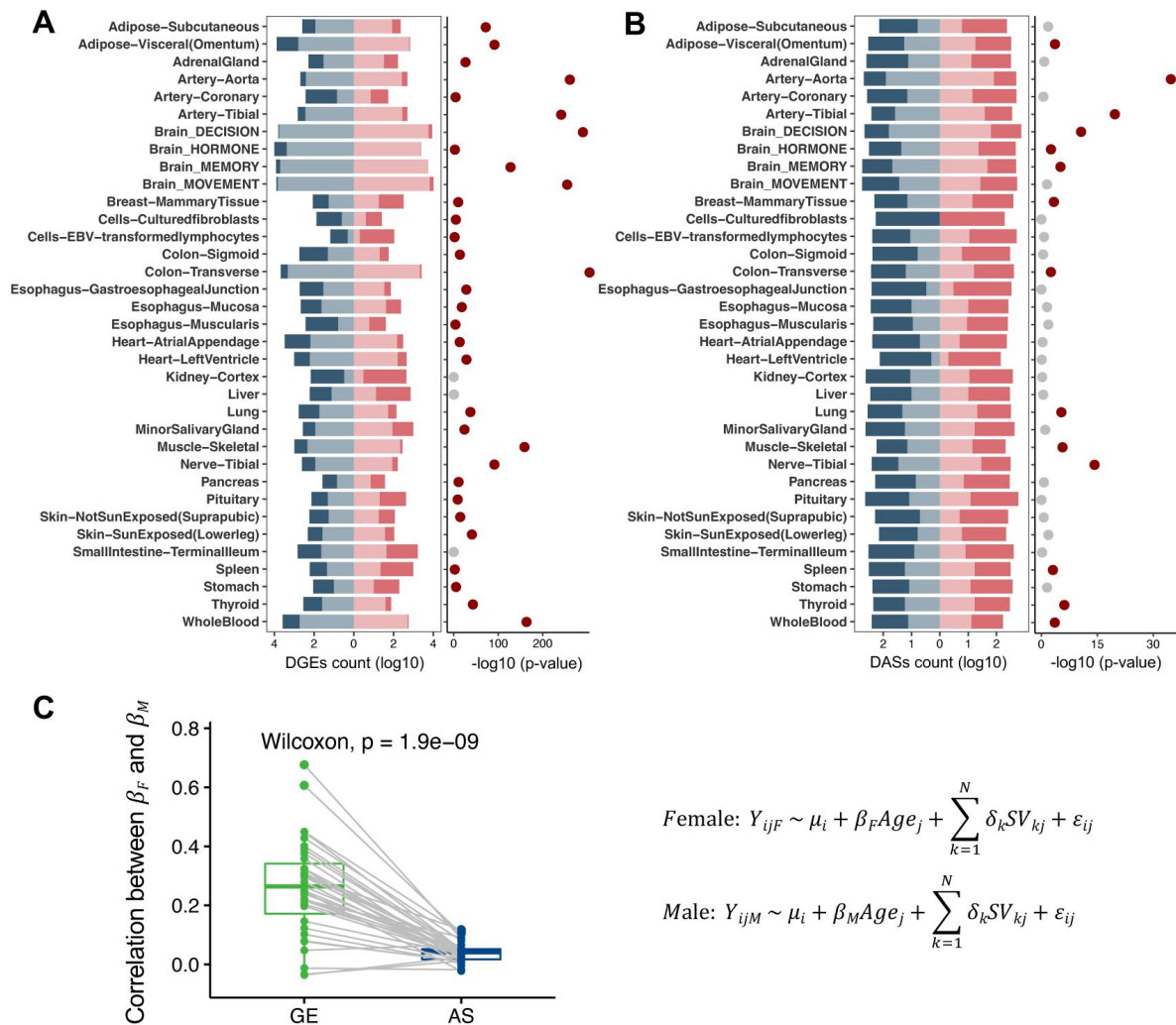


Fig. 3

Sex-stratified age-associated genes and AS events across tissues.

(A-B) The numbers of age-associated genes (**A**, left) and AS events (**B**, left) in each sex. The red bars indicate the numbers of age-associated genes or AS events in females, while blue bars indicate those in males. The numbers of age-associated genes and AS events common in both sexes are shown in bars with relatively lighter colors. The significance of the overlapped age-associated genes (**A**, right) or AS events (**B**, right) between sexes are estimated using hypergeometric test. Tissues with significant overlapped p-values are labeled by dark red dots.

(C) Correlation between the age effects on genes (green) or AS events (blue) in females vs. males. The Y-axis represents the correlations of the effect sizes of age (β_F , β_M) between two sexes across all genes/AS events (Spearman's correlation). Grey lines link a pair of tissue in GE and AS. The p-values are estimated using Wilcoxon signed-rank test.

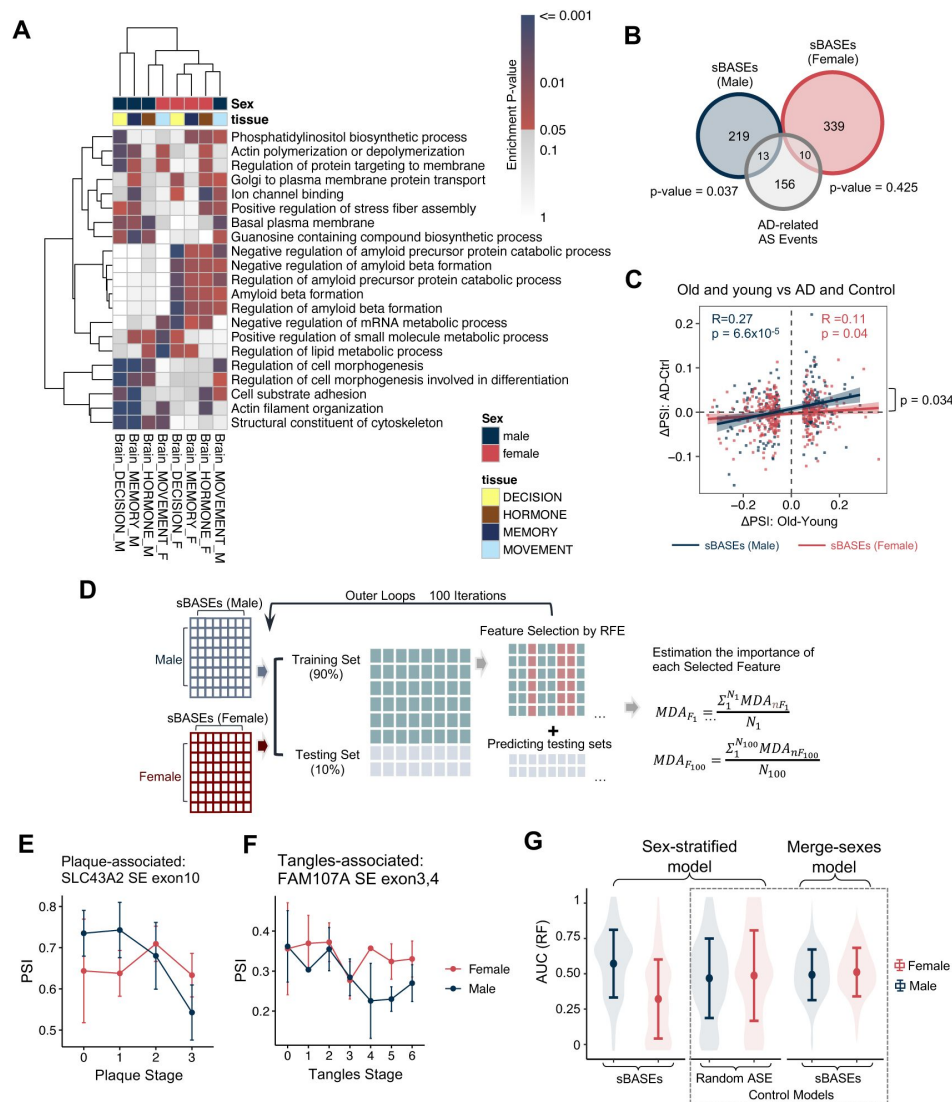


Fig. 4

Male-biased associations between the AS changes during aging and Alzheimer's disease.

(A) GO analysis of the sBASEs in each sex and brain region. Heatmap shows the sex-specific pathways that are significantly enriched in more than 3 brain regions. Clustering is conducted by default parameters in *pheatmap* functions. The $-\log_{10}$ transformed enrichment p-values are shown in the color scale.

(B) Venn diagram between the sBASEs in females and males with AD-related AS events. The p-values are calculated using hypergeometric test.

(C) Correlation between AD-associated and age-associated AS changes in females and males. The X-axis indicates the $\text{PSI}_{\text{Old}} - \text{PSI}_{\text{Young}}$, while the Y-axis indicates $\text{PSI}_{\text{AD}} - \text{PSI}_{\text{Control}}$. sBASEs in females are labeled in red, while sBASEs in males are in deep blue. The estimated Rho and p-value by spearman's correlation test in each sex are labeled on the top.

(D) Model for AD prediction and feature importance evaluation. sBASEs are used for predicting AD in females and males respectively. 90% of samples are randomly selected as training sets for 100 iterations. The recursive feature elimination approach is used for feature selection. Feature importance is evaluated by the averaged MDA across 100 iterations.

(E) AS levels of the skipped exon 10 on SLC43A2 during plaque stages in males and females.

(F) AS levels of the skipped exons 3 and 4 on FAM107A during tangles stages in males and females. Mean $\pm$ standard error is shown in the error bar.

(G) Performances of sex-stratified and merge-sexes models predicted by sBASEs or randomly selected AS events for 100 iterations. The control models (i.e., the sex-stratified model trained by randomly selected AS events and the merge-sexes model trained by sBASEs in each sex) are highlighted with dashed lines. The mean $\pm$ s.d. (standard deviation) of AUC across 100 iterations is shown in the boxplot.

PSI_{Ctrl}) was significantly correlated with the age-associated AS changes in males (i.e., $PSI_{Old} - PSI_{Young}$) (Pearson's product-moment correlation, $R=0.27$, $P=6.6 \times 10^{-5}$), and such correlation was significantly higher than that in a separate analysis in females using the female-specific sBASEs ($R=0.11$, $P=0.04$; **Fig. 4C** [↗](#); see Materials and Methods). In contrast, we found a significant overlap between the AD-related genes and the sex-biased age-associated GE changes in females ($p = 0.031$) but not in males ($p = 0.60$) (fig. S8B), supporting a recent finding that females are more vulnerable to AD during aging due to the post-menopause activation of C/EBP β -AEP/ δ -secretase pathway by sex hormone changes [44](#) [↗](#).

In addition, we performed similar analyses using an independent dataset from ROSMAP containing 271 filtered RNA-seq samples from dorsolateral prefrontal cortex (DLPFC) with matched age distribution between AD and control samples. As expected, we found that the AD-associated AS changes of sBASEs in males showed a significantly stronger positive correlation with the age-associated changes ($R=0.509$, $p < 2.2 \times 10^{-16}$) than such correlation across the sBASEs in females ($R=0.371$, $p < 2.2 \times 10^{-16}$) (fig. S8C), which is largely consistent with the earlier finding using a smaller dataset (**Fig. 4C** [↗](#)). This male-biased association indicated that the AS changes during aging could contribute more to the AD in males than females, suggesting additional molecular complexity in the sex effect of AD.

To further examine the contribution of individual sBASE to AD risk, we constructed a sex-stratified random forest classifier using a subset of sBASEs that were also detected in ROSMAP data for AD prediction. For each sex, we randomly divided the datasets of AD patients and controls into separate training and test sets, and repeated the prediction 100 times (**Fig. 4D** [↗](#)). For each selected sBASE, we also evaluated its averaged feature importance (i.e., mean decrease accuracy) across 100 iterations and the association with the classic AD neuropathology (such as tangles and plaques), resulting in the identification of a series of key sBASEs significantly associated with AD in a sex-specific manner. For example, the skipping of exon 10 in SLC43A2 negatively correlated with the neuritic plaques stage in males but not in females (**Fig. 4E** [↗](#)). Meanwhile, skipping of exons 3 and 4 in FAM107A, a gene related to synaptic and cognitive functions, showed a male-specific negative association with neurofibrillary tangle stages (**Fig. 4F** [↗](#)). In addition, the prediction models based on sBASEs performed significantly better in males than females (**Fig. 4G and fig** [↗](#). S8D, left panel), suggesting that age-associated AS events could serve as a better predictor for AD risk in male patients. As negative controls, we performed the same analysis to predict male and female patients separately using randomly selected AS events (sex-stratified model; **Fig. 4G** [↗](#), middle panel), and predict all patients with male- or female-specific sBASEs (merge-sexes model; **Fig. 4G** [↗](#), right panel). We found that both control models showed no differences in males and females, indicating the crucial roles of sBASEs in AD during male aging. Collectively, our analyses of human brains revealed a surprising sex-biased association between RNA splicing and AD, suggesting that regulation of AS may play a critical role in the sex-dimorphisms of age-related diseases such as AD.

sBASEs are regulated by splicing factors with sex-dimorphic expression during aging

AS is generally regulated by splicing regulatory *cis*-elements that specifically recruit various *trans*-acting splicing factors (SF) to promote or inhibit splicing of adjacent exons [27](#) [↗](#), [45](#) [↗](#). Dysfunction of the SFs has been reported as a hallmark of aging, presumably by regulating the genes in cellular senescence during aging process [33](#) [↗](#). Consistently, our analyses identified a large number of age-associated SFs in human brain (table S5). To further examine the sex-dimorphic regulation through age-associated SFs, we constructed an SF-RNA regulatory network between age-associated SFs and sBASEs in each sex. This process integrated the correlations between the SFs levels and PSI of sBASEs, the significant changes of sBASEs upon SFs knockdown, and the existence of SFs binding motifs around sBASEs (**Fig. 5A** [↗](#), left panel). We required the regulatory connections

between the SFs and AS events that pass all these three criteria. Such regulatory networks can be generated in different tissues, and we focused on the brain due to its association with age-related diseases such as cognitive disorders and synaptic diseases.

We identified several key age-associated SFs that regulate multiple sBASEs in decision-related brain region (**Fig. 5A** [↗](#), right panel). Interestingly, several SFs showed significant age association only in females (e.g., TIA1 in **Fig. 5B** [↗](#), left panel), while others were associated with age in both sexes (e.g., FXR2 in **Fig. 5B** [↗](#), right panel). This finding is consistent with our earlier observation that more sex-specific age-associated AS events were identified in decision-related brain region of females than males (**Fig. 4B** [↗](#)). Most of the sBASEs were regulated by sex-specific age-associated SFs rather than the SFs associated with aging in both sexes (**Fig. 5C** [↗](#)). Functional enrichment analysis showed that these sBASEs regulated by female-specific age-associated SFs were significantly enriched in glial cells and oligodendrocytes development (**Fig. 5D** [↗](#)), which participated in the myelin generation of aging brain and neurodegenerative diseases in central nervous system ⁴⁶[↗](#). These results indicated that the SFs with sex-dimorphic expression during aging, especially the female-specific changes of SFs, were one of the main factors in regulating sBASEs with important biological functions.

Furthermore, we studied the potential mechanism that controls the female-specific changes of SFs during aging in decision-related brain region. It was previously reported that the female sex hormone estrogen consistently decreases during lifespan ³⁷[↗](#),⁴⁷[↗](#); thus we speculated that estrogen may play roles in controlling transcription of SFs during aging through activation of nuclear receptors. To this end, we analyzed the public ChIP-Atlas data to examine the binding of canonical estrogen receptor ESR1 to the promoter region. We found that the ESR1 binding ratios on the female-specific age-associated SFs were significantly higher than the median value of the age-associated SFs common in both sexes (**Fig. 5E** [↗](#); $p = 0.032$), suggesting that ESR1 was more likely to bind the promoters of the female-specific age-associated SFs. In addition, the age-associated decline of ESR1 was more substantial in females than males (**Fig. 5F** [↗](#)), and some of the female-specific age-associated SFs (e.g., SRSF1 and SRSF7) could be stimulated by propyl pyrazole triol, an agonist of ER α (**Fig. 5G** [↗](#))⁴⁸[↗](#).

These observations were further confirmed by an independent experiment where the wildtype and two constitutively active mutants (Y537S, and D538G) of ESR1 were engineered into MCF-7 cells using CRISPR-Cas technology ⁴⁹[↗](#). The resulting cells were treated with estradiol or vehicle controls and subjected to transcriptome profiling by RNA-seq, and the data were reanalyzed using the same pipeline. We found that the basal levels of SRSF1 and SRSF7 were significantly higher in the cells with constitutively active ESR1 (**Fig. 5H** [↗](#), comparing samples treated with vehicle controls), and SRSF1 and SRSF7 were further induced by the estradiol treatment only in cells with wildtype ESR1 but not in cells with mutated ESR1 (**Fig. 5H** [↗](#), compare the samples treated with estradiol vs. control in different cells), indicating that the activation of ESR1 pathway can promote expression of SRSF1 and SRSF7. The similar changes were also found in other splicing factors including HNRNPA2B1 and HNRNPC (fig. S9A). As expected, the expression levels of SRSF1 and SRSF7 showed an obvious decrease during female aging (**Fig. 5I** [↗](#)).

To further confirm our findings in different tissues, we analyzed the RNA-seq data with ER knockout and estrogen treatments in mouse brain arcuate nucleus (ARCs) ⁵⁰[↗](#) and found a consistent trend in the expression changes for splicing factors Srsf1, Srsf7, and Hnrnpa2b1 under estradiol (E2) regulation (fig. S9B). Specifically, these splicing factors were activated by estrogen in wild-type mice but exhibited reduced expression in ER knock-out mice. The direct binding of estrogen was also examined through integrating the ESR1 ChIP-seq data ⁵¹[↗](#). As expected, a major ESR1 binding peak was found in the promoter regions of Srsf7, which was increased by estrogen treatment (fig. S9C).

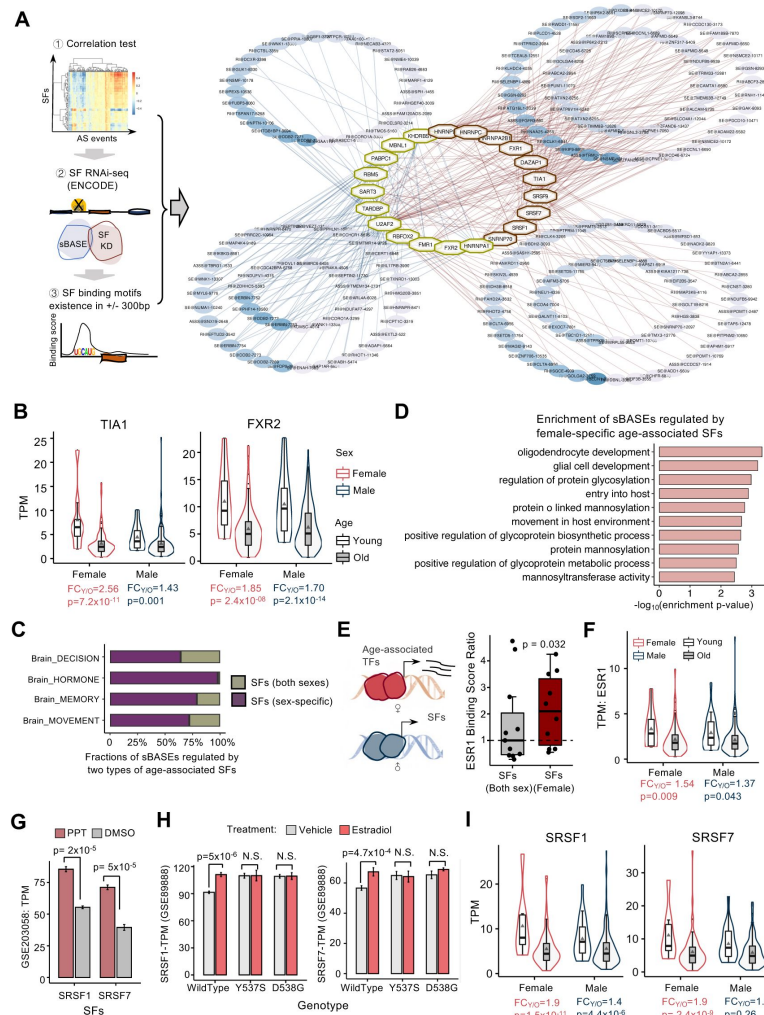


Fig. 5

Sex-dimorphic AS regulation during aging in decision-related brain region.

(A) Schematic diagram for constructing AS regulatory networks (left) and the networks of the decision-related brain region (right). The blue and red lines indicate the regulations in males and females. Hub genes in octagons are SFs. The red ones are female-specific age-associated SFs, while the age-associated SFs common in both sexes are labeled in yellow. The color of the ellipse indicates the significance of sBASEs during aging, and the thickness of the line shows $-\log_{10}$ spearman's correlation p-values between the TPMs of SFs and PSIs of AS events during aging.

(B) Examples of female-specific age-associated SFs (left) and age-associated SFs common in both sexes (right) during aging. The p-value and foldchange between old and young are labeled at the bottom.

(C) Percentages of the sBASEs regulated by the sex-specific age-associated SFs in four brain regions.

(D) Functional enrichment of sBASEs regulated by female-specific age-associated SFs based on MsigDB.

(E) Schematic diagram for the transcriptional regulation on age-associated SFs via nuclear receptors (left). Boxplot shows the ESR1 binding scores on age-associated SFs (right). The Y-axis indicates the ratio of ESR1 binding scores of each SF divided by the median binding score of age-associated SFs common in both sexes.

(F) Expression levels of ESR1 during aging. The legends of each group are the same as those in **B**.

(G) Expression levels of SRSF1 and SRSF7 treated by estrogen receptor agonist (PPT) vs. DMSO in MCF-7 cell line (N=3). The error bars indicate the mean $\pm$ s.d..

(H) Expression levels of SRSF1 and SRSF7 treated by 1nM estradiol (E2) vs. vehicle control (veh) in ESR1 wildtype, Y537S, and D538G mutant MCF-7 cells (N=4).

(I) Expression levels of SRSF1 and SRSF7 during aging. Y-axis indicates the normalized TPM values. The p-value and foldchange between old and young groups in each sex are labeled at the bottom. The legends of each group are the same as those in **B**.

Collectively these analyses across multiple datatypes suggested that the female-specific SFs (such as SRSF1 and SRSF7) are more sensitive to the estrogen changes *via* ESR1-mediated pathway, leading to extended sex-biased AS regulations during aging.

The aging rates of GE show sex-dimorphism during aging across multiple tissues

In the earlier results (Fig. 3A), we found a large proportion of age-associated genes are common in both sexes. Therefore, we asked whether the levels of GE change at the same pace during aging in females and males. The age-associated genes ubiquitously changed in both sexes were analyzed for consistency. We first used the Autoregressive Integrated Moving Average model⁵² to capture the age-associated genes with significant chronological changes beyond random fluctuations, which are the main contributors to transcriptome dynamics during aging. We further quantified the GE changes between adjacent age windows using breakpoint analysis¹⁶ (Fig. 6A, see Materials and Methods). Such analyses were conducted across all tissues in both sexes (fig. S10A), with two tissues shown in Fig. 6B. A higher value on the Y-axis (mean -log P) indicated a larger GE change at the given age, and thus the peaks were regarded as aging “breakpoints” that represent the key time point of transcriptome aging. Compared to females, males showed significantly larger GE changes during aging in most tissues (fig. S10A), except transverse colon, fibroblast cells, and whole blood (fig. S10B). The number of breakpoints also showed sex-dimorphism in several tissues. For example, in sun-exposed skin and colon, males showed two main breakpoints around ages 35 and 50, whereas females only have one breakpoint at age ~45 (fig. S10A).

Moreover, we found that the major aging breakpoints (i.e., breakpoints with the largest GE changes) occurred earlier in males than females across most tissues, and the magnitudes of changes were also more prominent in males (Fig. 6C), suggesting that males may age faster with a larger age-associated GE change. In addition, we conducted breakpoint analysis with all chronological genes in both sexes and found similar results when comparing the breakpoints in males and females (fig. S10C). We also performed a similar analysis using chronological sex-biased age-associated AS events (sBASEs) and all chronological AS events in females and males separately, and again found an earlier and more obvious breakpoint (fig. S10, D and E). In summary, these results uncovered an earlier aging and a higher aging rate in males at molecular levels, again implying that males could be surprisingly more vulnerable to aging.

Aging contributors in separate sexes show distinct functions

To further examine how the individual gene affected aging rates, we designed a disturbance analysis to identify aging-modulated genes (AMGs) (Fig. 6D). Specifically, we randomly removed 20% of chronological genes and repeated the breakpoint analysis with the remaining genes to test if the breakpoints were significantly altered (see Materials and Methods). This process was repeated 200 times, and the genes enriched in the 200 random samplings (judged by Fisher exact test $p < 0.05$) were defined as AMGs that serve as the key contributors to molecular aging.

We identified a total of 1035 AMGs across different tissues (table S6), around 70% of which (723 in 1035) were sex-specific (Fig. 6E). Functional enrichment analysis suggested that the female-specific AMGs were significantly enriched in cellular respiration and NADH dehydrogenase pathways (Fig. 6F), which is consistent with an earlier report of their roles in human aging and longevity⁵³. Such functional enrichment was not found in the male-specific AMGs (Fig. 6G), implying that female aging may be more sensitive to mitochondrial and oxidative functions. Intriguingly, we also found many male-specific AMGs relevant to myeloid leukocyte differentiation, phagocytosis, and macrophage apoptotic processes, implying a sex-differential role of the immune system during aging.

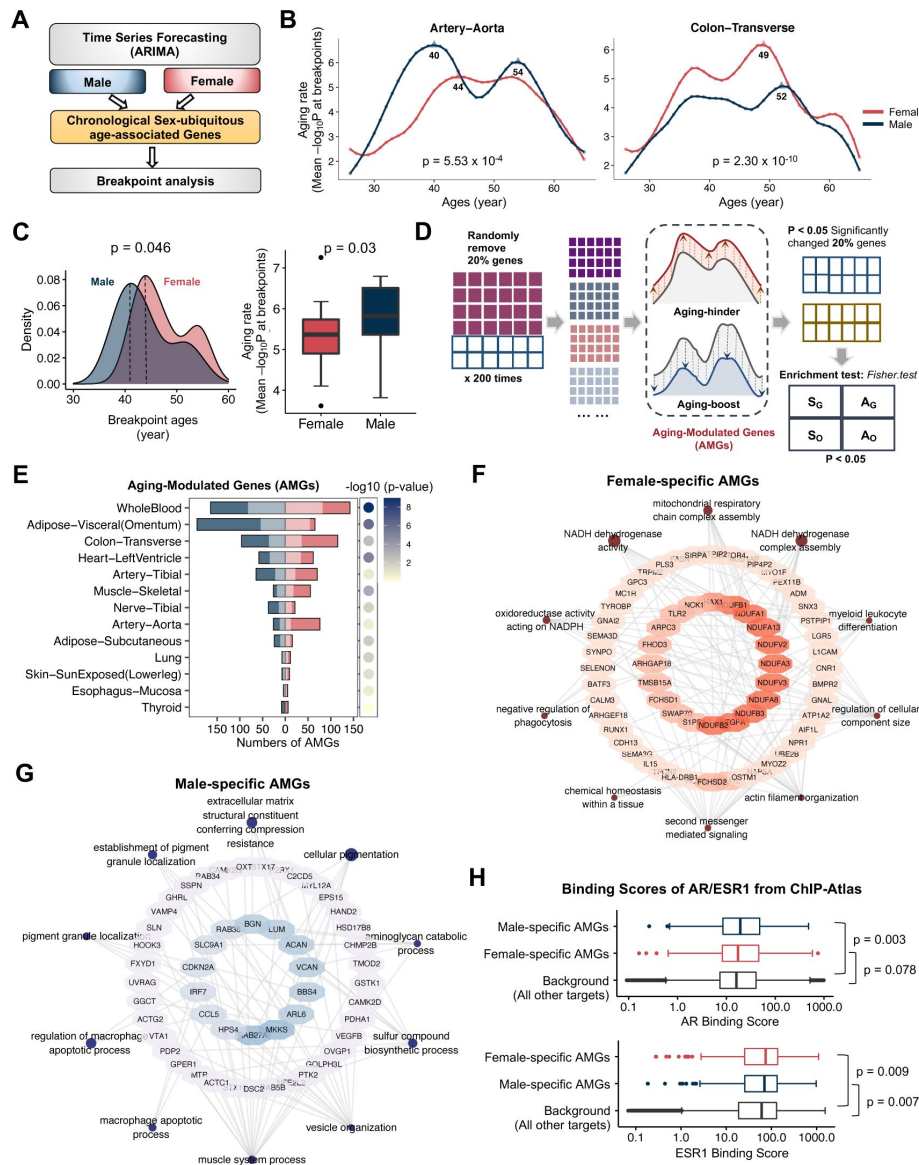


Fig. 6

Sex-dimorphic aging rate of GE during aging process.

- (A) Workflow of time series and breakpoint analysis of gene expression in each sex.
- (B) Aging rates and breakpoints in the aorta (left) and transverse colon (right). The Y-axis shows the estimated aging rate (see Materials and Methods). The p-values are calculated using Wilcoxon signed-rank test.
- (C) Characteristics of the major breakpoints across tissues in each sex. The left panel shows the distribution of the major breakpoints across multiple tissues. The right panel shows the aging rate at the major breakpoints. The p-values are calculated using Wilcoxon signed-rank test.
- (D) Pipelines for identifying AMGs in each sex and tissue (see Method).
- (E) AMGs in different sexes and tissues. The numbers of AMGs in females (red) and males (blue) are shown in the barplot (left). Tissues are ordered by the number of AMGs common in both sexes, which are shown in lighter colors. Significances of the overlapped AMGs are calculated using hypergeometric test and colored in a gradient palette (right).
- (F-G) GO analysis of male-biased AMGs (F) and female-biased AMGs (G). Sex-specific GO terms with top10 $-\log_{10}$ transformed enrichment p-values are plotted. GO terms are labeled by purple dots, the size of which indicates the significance of the enrichment analysis. The larger the dots, the more significant enrichment scores of the biological pathways. AMGs are labeled by blue dots, the color of which indicates the number of GO terms associated with those genes.
- (H) The binding scores of AR (top) and ESR1 (bottom) on male-specific AMGs, female-specific AMGs, and randomly selected AR/ESR1 target genes. The p-values are calculated by Wilcoxon rank-sum test.

To determine the potential mechanism of the sex-dimorphic aging rate, we focused on sex hormones that are critical for transcriptional regulation. We found that the androgen receptor (AR) tends to bind the promoters of male-specific AMGs ($P=0.007$), suggesting potential regulations by androgen during male aging (**Fig. 6H** [↗](#), top panel). On the other hand, female-specific AMGs were more likely to be bound and regulated by estrogen receptor ESR1 (**Fig. 6H** [↗](#), bottom panel), suggesting the potential regulation by distinct sex hormones.

Furthermore, we examined the detailed expression profile of AMGs from the whole blood that is commonly used for health surveillance. The AMGs common to both sexes can be grouped into three clusters based on age-associated expression profiles. GE of the largest cluster increased at young ages and abruptly decreased in mid-ages in males, but monotonically decreased in females (fig. S11, A and B). These genes were functionally enriched in neutrophil and granulocyte migration and RNA splicing (fig. S11C). Moreover, we also examined sex-specific AMGs and found very distinct expression patterns (fig. S11, D and E). Interestingly, while most of these genes showed decreased expression at old ages, the expression of some male-specific AMGs gradually increased during aging (fig. S11D, dotted box). Taken together, these results suggested that these age-associated genes modulating the transcriptome aging rate showed different functional impacts in females and males.

Discussion

Previous studies in selected tissues suggested that sex is a key factor in affecting human aging and age-related diseases ^{11,12}, which may contribute to different life expectancies between two sexes. However, few pan-tissue investigation on age-by-sex effect was conducted at the transcriptome level, especially in the resolution of AS isoforms. By comprehensively analyzing ~17,000 samples across 54 tissues, we uncovered individual sex or age effect on global transcriptome variations, as well as the combined age-by-sex effects. Using a sex-stratified differential analysis, we found that AS changes showed a stronger discrepancy between females and males during aging. We further identified the regulatory splicing factors for sBASEs and explored the underlying mechanism of sex-differential AS regulation during aging. Conversely, we found the change of age-associated gene expression was consistent between two sexes, with a faster aging rate in males. Our analysis provided a comprehensive landscape of how sex and age affect human transcriptome, revealing the sex-dimorphic aging at a multi-tissue level.

Multiple mRNA isoforms can be produced from a single gene through AS, increasing the transcriptomic and proteomic complexity in different cells ^{54,55}. Therefore, analyses of AS changes during aging of each sex may provide additional information beyond the GE profile. Interestingly, we observed different patterns when analyzing the profiles of GE vs. AS. For example, judged by the AS changes, we observed more tissues with a significant sex or age effect on transcriptome and a higher discrepancy between females and males from the combinatorial effects of age-by-sex (**Fig. 3** [↗](#)), highlighting the contribution of splicing in sex dimorphism during human aging at an isoform-specific resolution.

The sex-dimorphic patterns of AS and GE not only illuminate aging-related transcriptomic changes but also reveal intriguing connections to age-related diseases, such as Alzheimer's disease (AD). Most studies showed that females have a higher risk of AD due to the post-menopause changes of sex hormones ⁴⁴, however, a higher incidence rate of mild cognitive impairment in males is also reported ⁵⁶. Additionally, based on an early study on a large cohort of 17,127 participants from UK Biobank, age-related cognitive risk factors are more widely discovered in males, suggesting a potential vulnerability of males to cognitive decline with age ⁵⁷. We further validated our findings using Alzheimer's disease dataset, ROSMAP, where the consistent correlations between aging-related splicing changes and AD-related changes were observed, providing additional evidence for the robustness of our results. At the splicing level, we surprisingly observed a male-

biased association between age-associated AS changes and AD (**Fig. 4**). Such associations could not be detected by analyzing GE, suggesting that analysis at isoform resolution can provide different information into sex-dimorphic aging. We also designed a sex-stratified random forest classifier to predict AD in males and identify several key AS events contributing to such prediction. Due to the limited sample size, large variations were observed between different training/testing iterations during the sex-stratified AD prediction. However, we still observed a higher performance of the classifier in males compared to the control model, which is in concordance with that the AS changes during aging may contribute more to AD in males than females. Our findings suggest that male-biased age-associated AS changes may contribute to the increased vulnerability of males to cognitive decline, providing a complementary perspective to existing research. We expect that future investigations using bigger datasets will further identify the sex differences in AD.

Using a pan-tissues analysis, we discovered that transcriptomic aging happened earlier in males than in females (**Fig. 6D**), consistent with the reported shorter life expectancy in males. Furthermore, we identified many sex-specific AMGs with different biological functions, such as the macrophage and mitochondria functions (**Fig. 6**, F and G). These genes tend to be regulated by the dominant sex hormones in each sex. The investigation of the aging rate provided insights into the intervention of the aging process or age-related diseases, and to some extent made available for identifying therapeutic targets at specific age points. It is worth noting that our analysis of the aging rate in this section does not include the data from brain tissues due to the sample sparsity in selecting parameters.

Our analyses also have some technical limitations, the biggest of which may be caused by the cell type and sample heterogeneity between different donors of GTEx datasets. Since GTEx dataset is the most comprehensive dataset with both transcriptome and whole genome sequencing data, it's a good public resource for sex, age, and age-by-sex interaction analysis within a healthy context. Compared with the earlier studies that used either microarray or other RNA-seq data to analyze the effect of age or sex on brain transcriptome, our main findings supported the main conclusion of these reports, including the diverse gene profiles across brain regions during aging, as well as more vulnerability in males on a global scale. In addition, GTEx has a very strict selection of healthy donors to reduce biases from different diseases, which is a unique feature compared to other cohorts that often contain patients of various diseases that may affect the analyses at the molecular level. In particular, the large-scale brain samples provide a unique opportunity to analyze transcriptomic changes in sex-dimorphic aging. However, several technical challenges could limit the generalizability due to dataset-specific biases, including cell type heterogeneity, postmortem artifacts, as well as sequencing biases. For example, GTEx data is bulk RNA-seq, which does not capture cell-type-specific transcriptomic changes. Given the cellular complexity of the brain tissues, the observed differences in gene expression and splicing may be influenced by shifts in cellular composition rather than intrinsic transcriptional regulation.

The second limitation is the existence of potential confounding factors (like batch effect, genetic background, etc.) that are often associated with big data analyses. However, when such batch effects are entangled with the effect from distinct sex/age groups, it becomes technically challenging to effectively remove such factors without canceling out the contribution of sex/age, which is a common technical difficulty reported also in other analyses. We take several measurements to reduce the effect of confounding factors. First, we utilized the original TPM/PSI data to aggregate sex/age effects and conducted a permutation analysis to assign an empirical *p*-value to each pcSVR. This approach provided a robust statistical framework for evaluating the significance of sex/age effects while accounting for potential batch effects. Secondly, in the context of the differential analysis, we preserved all possible confounders and identified changes that were more obvious than the effects of these confounders, a strategy that reduces false positive discoveries originating from potential batch effects. Another technical limitation is related to the

small number of samples in some tissues, and thus we either omitted them from aging analyses (like pituitary, bladder, and kidney) or combined similar tissues (like brain regions) for detailed analyses.

Moreover, we derived the AS regulatory networks using the multi-dimensional large-scale data from highly proliferative and easily cultivable cancer cell lines, which may not be an ideal model for the post-mitotic cells in human tissues. However, we dissected the RBP-RNA regulatory network by integrating multi-dimensional data obtained through several orthogonal state-of-the-art approaches (**Fig. 5A**), including the eCLIP-seq, *in vitro* RNA affinity evaluation, RNA-seq with RBP depletion, and chromatin association⁶⁰. In addition, the RBP-RNA binding relationships seem to be universal, and the distinct splicing outcomes are mainly determined by the tissue-specific expression and activity of different splicing regulatory factors. Therefore, it is possible to use the data in cancer cells to derive a connectivity map of RBP-RNA, and then use the tissue-specific expression of RBPs to refine the regulatory map.

It was reported that there is only a moderate correlation between mRNA and protein abundances for many human genes, mainly because of the variation in translation efficiency of different mRNAs and the additional regulation in protein degradation. Here we constructed the SF-RNA regulation network with the shRNA-seq data that represents the reduction of specific splicing factors in both RNA and protein levels. Such data is more reliable than the correlation of cell-specific expression of RNA and protein, and thus it is feasible to evaluate the consequences of gene silencing by looking at the RNA level. To minimize tissue biases in the RBP-RNA regulation network, we incorporated 3-step processes (see Materials and Methods) including the restrictions of the correlations between RBP expression and AS events within each tissue. Additional large-scale experiments are needed in the future to globally profile the physical and functional interactions between RNA and RBP with a higher accuracy from different tissues.

Mechanistically, we found a prominent role of sex hormones in regulating sex dimorphic aging or age-related diseases. Using a range of RNA-seq and ChIP-seq data, we uncovered a type of female-specific age-associated AS regulation (**Fig. 5A**), which may be achieved partially through the estrogen-mediated regulation of female-specific splicing factors (**Fig. 5**, E and I). We further quantified chronological changes of transcriptome dynamics using breakpoint analysis, and observed the sex-dimorphic aging rates across multiple tissues. Our analysis observed the non-linear aging patterns with two breakpoints, which is consistent with recent findings, with some differences in specific age points due to sex differences as well as tissue diversities⁶¹. In addition, our aging breakpoints are largely consistent with the changes in circulating sex hormones during aging. The level of testosterone starts to decrease at the age 35–40 in males, while in females, estrogen level starts to decrease at a later age (i.e., ~50 years old)^{37,47}. These breakpoints could represent key junctures in the aging process that align with the non-linear patterns of aging and disease progression. Furthermore, we identified a series of female-specific and male-specific aging-modulated genes (AMGs), whose functions are consistent with the reported roles of sex hormones. For example, estrogen was reported to participate in the antioxidant system via NADPH^{37,47}, which is functionally enriched in female-specific AMGs (**Fig. 6F**). Conversely, androgen was reported to affect the inflammation in immune response^{37,47}, which is enriched in the male-specific AMGs (**Fig. 6G**). Consistently, we found that the sex-specific AMGs were more likely to be regulated by androgen receptor in males or estrogen receptor in females (**Fig. 6H**). However, the regulation of sex-biased aging is probably more complicated. Other factors, including mitochondrial functions, genes encoded by sex chromosomes, non-coding RNAs, and transposable elements, may also contribute to the sex-dimorphism during aging^{12,14,62,63}. In addition, it is possible that the genetic variants, including expression QTL (eQTLs) and splicing QTLs (sQTL), can affect GE or AS in a sex-biased or age-associated fashion^{22,40,64}. Future studies of these factors will improve our understanding of the intricate regulation of transcriptome dynamics, providing us a broader horizon at a multi-omics scale.

Various biological aging clocks had previously been constructed based on different molecular markers, including DNA methylome ⁶⁵[65](#)–⁶⁷[67](#), transcriptome ⁶⁸[68](#), circulating immune proteins ⁶⁹[69](#), human gut microbiome ⁷⁰[70](#), and multi-omics features ⁷¹[71](#). Such aging clocks generally pay little attention to the difference between females and males. However, our results advocate for a sex-specific age model. In addition, certain age-related diseases may accelerate aging process, as tumors aged 40% faster than matched normal tissues based on DNA methylome ⁷²[72](#). Therefore, we speculate that constructing separate aging clocks in distinct sexes will help model the progression of age-related diseases accurately.

Materials and Methods

Transcriptome quantifications and data processing

We used the transcriptome data from Genotype-Tissues Expression (GTEx) project version 8.0, which contains 17,382 samples in 54 tissues (including 13 brain regions and two cell lines). We downloaded the RNA-seq bam files and the phenotype information from dbGaP (study accession: phs000424.v8.p2; table accession: pht002743.v8.p2.c1). The tissues with a small number of samples and specific to only a single sex (such as ovary and testis) were removed for a better comparison. In addition, for each donor, we combined 13 brain regions into 4 main functional regions, including hormone- or emotion-related region (amygdala, anterior cingulate cortex, and hypothalamus), movement-related region (cerebellar Hemisphere, cerebellum, and spinal cord), memory-related region (caudate, hippocampus, nucleus accumbens, putamen, and substantia nigra) and decision-related region (cortex, and frontal cortex). As a result, for the N donors that each have 13 brain regions (N×13 sample matrix), we merged the data into 4 functional regions to generate an N×4 sample matrix in the downstream analyses. In total, 16,202 samples in 35 tissues or brain regions (31 tissues and 13 brain regions) from 948 donors passed these filters for further analysis. The numbers of samples in each tissue are shown in table S1. GE and AS quantifications were performed by Paean ⁷³[73](#), a parallel computing system on the GPU-CPU platform with high computational efficiency for super-large datasets. We calculated TPM (transcripts per million) for protein-coding genes and PSI (percent of spliced-in) values for AS events, including skipped exon (SE), alternative 5' splice site (A5SS), alternative 3' splice site (A3SS), and retained intron (RI). Genes with low expression (average TPM < 1) were removed. We also filtered the AS events by first removing the samples with missing PSI values in more than 50% of AS events (16,182 samples left), and then filtered the AS events using the following criteria: (1) percentage of the missing PSI values in < 5% samples; (2) averaged total reads counts in spliced junctions (i.e., spliced in counts + spliced out counts) > 10; (3) non-constant PSI values across samples; (4) max(PSI) – min(PSI) > 0.05; (5) standard deviation > 0.01; (6) averaged PSI value in the range from 0.05 to 0.95; (7) TPM of the spliced genes > 1. We focused on the AS events in protein-coding genes. In different tissues, the average numbers of genes and AS events for downstream analysis were 12,144 and 13,939, respectively.

Even after removing the AS events with too many missing values, the remaining AS data were still too sparse in some samples for a reliable statistical analysis. On average there are about 0.17% AS events missing from the RNA-seq samples in different tissues. For a single RNA-seq sample in certain tissues, there were up to 15% AS events still missing (i.e., below the detecting limit). Several imputation techniques (reference-based and reference-free) have been widely used and easily implemented in multi-omics datasets ⁷⁴[74](#), and we used the k-nearest neighbors (KNN) algorithm for missing PSI value imputation in this analysis.

Principal component-based signal-to-variation ratio (pcSVR)

To evaluate the sex/age effect on human transcriptome, we designed a principal component-based signal-to-variation ratio (pcSVR) based on principal component analysis (PCA) and signal-to-noise ratio (SNR). PCA process could maximize the global transcriptomic variations, and SNR could

calculate the distances between sex/age groups divided by the noises within each group (Equations 1-3, female vs. male and young vs. old). We selected the principal components that captured more than 80% of the global variation (other PC cutoffs capturing 50%-90% of variations were also tested as controls). The SNR of the principal component score was calculated according to the following formula:

$$pcSVR(X, Y) = \frac{\|\bar{X} - \bar{Y}\|_2}{\sqrt{\frac{\sigma_X^2}{N_X} + \frac{\sigma_Y^2}{N_Y}}} \quad (1)$$

$$\sigma_X^2 = \frac{\sum_{i=1}^{N_X} \|x_i - \bar{X}\|_2^2}{N_X - 1} \quad (2)$$

$$\sigma_Y^2 = \frac{\sum_{i=1}^{N_Y} \|y_i - \bar{Y}\|_2^2}{N_Y - 1} \quad (3)$$

where X and Y label two groups with sample sizes N_X and N_Y (females and males; young and old); x_i and y_j are the PC scores in sample i ; and $\bar{X}$ and $\bar{Y}$ are average PC scores across samples in each group calculated by GE or AS.

Here, to figure out whether the pcSVR is statistically significant enough, we conducted the permutation analysis and derived empirical p-values of the pcSVR. We first randomly selected subsamples (50% numbers of the minimum sample sizes of two groups) 10,000 times in each sex/age group and computed the averaged pcSVR. Then, we also randomly selected the same number of subsamples 10,000 times in each tissue regardless of the sex/age labels. The empirical p-values were calculated by the percentage of tested 10,000 pcSVR values (i.e., calculated using 50% sub-sampling irrespective of the group labels for 10,000 repeats) higher than the average pcSVR (i.e., calculated using 50% sub-sampling with the group labels for 10,000 repeats). The null hypothesis is that the two sets of pcSVR values are drawn from the same distribution. We performed all of the random sampling approaches using even probabilities in the R function ‘sample’ with the parameter prob = NULL. The tissues with more than 3 samples in each age group in females or males are included for this analysis (total 29 tissues and 4 brain regions).

To compare different approaches for empirical p-value estimation, a bootstrap with the replacement approach was also performed for null distributions. The process was repeated 10,000 times to minimize the biases of random sampling to construct a proper null distribution. Compared with the boot-strapping approach with replacement, the sub-sampling approach could also evaluate the robustness of pcSVR values to the sample size. Genes and AS events encoded by the protein-coding genes with averaged TPM > 1 were considered. To evaluate the contribution of sex chromosomes to pcSVR, we also removed the genes and AS events encoded by sex chromosomes and recalculated the pcSVR with corresponding p-values in each tissue.

Identification of differentially expressed genes and AS events

A recent study reported that DESeq2 and edgeR have unexpectedly high false discovery rate (FDR) in analyzing large-scale datasets, as well as the inconsistency between differentially expressed genes identified by these two approaches⁴³. Linear regression models have been widely used for transcriptomic and genomic differential analysis in recent studies^{22,32,35,75}, which have the advantages of controlling all known (e.g., sex, age, race, etc.) and unknown variables (e.g., confounders and other batch effects, etc.). In each tissue, we fitted a linear regression model including age, sex, and age-by-sex interactions as covariates, as well as other confounding factors. The following linear regression was used:

$$PSI_{ij} \sim \mu_i + \alpha SEX_j + \beta AGE_j + \gamma SEX_j \times AGE_j + \sum_{k=1}^N \delta_k SV_{kj} + \varepsilon_{ij} \quad (4)$$

$$TPM_{ij} \sim \mu_i + \alpha SEX_j + \beta AGE_j + \gamma SEX_j \times AGE_j + \sum_{k=1}^N \delta_k SV_{kj} + \varepsilon_{ij} \quad (5)$$

where PSI_{ij} is the normalized PSI value for AS event i in sample j ; α , β , γ are the coefficients of sex, age, and age-by-sex interactions, respectively; μ_i is the regression intercept; ε_{ij} is the error term; and k is the number of surrogate variables (SVs) in N total SVs for sample j with the coefficients δ_k . SVs are estimated using SVA (Surrogate Variable Analysis)⁷⁶. For differential GE analysis, we used the same linear regression model for analyzing differential GE by replacing PSI with TPM in the formula (Equation 5). To deal with the different orders of magnitude between PSI and TPM values, we normalized the PSI/TPM of each gene/AS event across samples using a scaling method to normalize the distance from the mean PSI/TPM across all samples.

$$norm(Y_{ij}) = \frac{Y_{ij} - \text{mean}(Y_j)}{\text{AveDis}(Y_j)}; \text{AveDis}(Y_j) = \frac{\sum_{i=1}^n |Y_{ij} - \text{mean}(Y_j)|}{n}$$

Here, Y_{ij} indicates the PSI or TPM of the gene/AS event i in sample j , and the $\text{AveDis}(Y_j)$ indicates the average distance from the mean PSI/TPM across all samples.

We defined sex-differential or age-differential genes with the cutoffs of p-value < 0.05 for the α or β coefficients, and foldchange > 1.5. Sex-differential or age-differential AS events were defined with the cutoffs of p-value < 0.05 for the α or β coefficients, and $|\Delta\text{PSI}| > 0.05$. We performed the differential analysis in each gene or AS event independently. The γ significantly larger or smaller than 0 indicates significant age-by-sex interaction, which suggests that the values of TPM and PSI are not only determined by sex or age independently but also affected by the age at a specific sex or sex at a specific age.

To evaluate whether the genetic effects could be controlled by SVs, we calculated the correlation between SVs and donors' ethnicity using Point Biserial Correlation which measures the association between continuous variables and dichotomous variables. Also, the Spearman's correlation was used to calculate the correlation between SVs and the top5 principal components as judged by whole genome sequencing data (WGS). The ethnicity information was extracted from phenotype data. The results of principal components calculated by the WGS dataset were downloaded from GTEx portal.

Permutation analysis was used to address the false discovery issue of this linear regression model in the differential analysis. We first generated the permuted data by randomly shuffling the group labels (sex or age) with the same sample sizes as the original group labels. The differential genes/AS events (DEGs/DASs) were identified based on each generated permuted phenotype data. This process was performed 1,000 times and we used two strategies to examine the robustness of our cutoffs. We first defined the fractions of original DEGs/DASs which are also identified in each permutation iteration. Meanwhile, the FDR is defined as how many permutations incorrectly identified the specific DEG/DAS across 1,000 permutations.

Evaluation of the effect of the sample sizes in differential GE and AS analysis

The effects of GTEx sample sizes on the numbers of differential genes/AS events were evaluated by a linear regression model by the following formula:

$$N_i \sim \mu + \alpha \text{Size}_i + \beta \text{Detected}_i + \varepsilon \quad (6)$$

where N is the number of sex-differential or age-differential genes/AS events in tissue i ; Size is the number of selected samples; Detected is the detected numbers of genes/AS events after filtering in preprocessing procedures; α , β are the coefficients of the sample sizes and numbers of detected genes/AS events, respectively; μ is the regression intercept; and ε is the error term.

Identification of sex-stratified age-associated genes/AS events

To identify the genes/AS events affected by age in separate sexes, we also conducted differential analysis by fitting the following linear regression model in females and males, respectively:

$$Y_{ij} \sim \beta_0 + \beta_{F/M} AGE_j + \sum_{k=1}^N \delta_k SV_{kj} + \varepsilon_{ij} \quad (7)$$

where Y_{ij} is the normalized TPM or PSI values of gene/AS event i in sample j ; $\beta_{F/M}$ is the coefficient of age in females or males; β_0 is the regression intercept; ε_{ij} is the error term; and k is the number of SVs in N total SVs for sample j with the coefficients δ_k evaluated in each sex. Age-associated genes or AS events were defined according to a significant age effect in different sexes (β_M , β_F , with cutoffs of $p < 0.05$; GE fold change > 1.5 or AS $|\Delta PSI| > 0.05$). The biases in different sexes also reflected the functional interactions between sex and age. This approach not only could indicate whether there is a sex-by-age interaction, but also define the direction of sex-by-age interaction, that is, how females/males affect GE/AS during aging differently. In addition, ΔPSI and Foldchange have also been used for filtering, which are the important thresholds to evaluate the degree of changes. Thus, we further defined the female- or male-specific age-associated genes/AS events using this sex-stratified analysis, and classified these age-associated genes/AS events into sex-biased age-associated genes/AS events (i.e., age-associated genes or AS events specific to one sex) or sex-ubiquitous age-associated genes/AS events (i.e., age-associated genes or AS events in both sexes).

Functional enrichment and Disease Ontology analysis

Potential functions of sBASEs were expounded based on C5 collections (ontology gene sets) in Molecular Signatures Database (MSigDB) and Disease Ontology. We carried out these processes using Bioconductor R packages *msigdb*⁷⁷, *clusterProfiler*⁷⁸, and *DOSE*⁷⁹, with the significance threshold p -value < 0.05 . The GO terms with the significances specific in only one sex per tissue were selected for downstream analysis. In the enrichment heatmap of brain regions (Fig. 4A), we plotted the GO terms with significance in at least 3 brain regions. Because the enrichment heatmap across all tissues are unrealistically large, we plotted the terms within at least 15 tissues and the tissues/sex with significance in at least 9 significant GO terms (fig. S6). The entire results of the enrichment analysis were shown in table S4.

Processing of Alzheimer's disease datasets and construction of AD prediction model

The AD samples of human prefrontal cortex were selected from GEO datasets (ID: GSE174367) containing 375 samples of single-nucleus RNA-seq, single-nucleus ATAC-seq, and bulk RNA-seq. We filtered the datasets by selecting the RNA-seq samples from the Frontal Cortex brain region (Library Source: 'TRANSCRIPTOMIC'; Brain region: 'FC'). Next, we selected the samples with the isolation approach 'total RNA isolation'. Due to the sequencing depth on the splicing junction, we filtered out the samples lower than 1Gb Bytes (SRR14514282, SRR14514349, SRR14514308, and SRR14514275). The remaining 91 samples were selected for GE and AS quantification, including 45 AD samples and 46 control samples from 47 male and 44 female donors (Female:Male=1:1; AD:Control=1:1) with similar age distribution between AD and control group. Quality control and adaptor trimming were performed using FastQC⁸⁰ (version 0.11.9) and Trim Galore (version 0.6.6). Bowtie2⁸¹ (version 2.4.2) was used to remove reads mapping to ribosomal RNA. Then, clean reads were mapped to human reference (GRCh38) using STAR⁸² (version 2.7.6). We also used pre-mapped RNA-seq data from ROSMAP containing the RNA-seq samples from the bulk sequencing of DLPFC. We further filtered the data to match the number of both sexes and the age distribution between AD and control groups, resulting the 215 AD and 56 control samples. The quantifications of GE and AS were performed by Paean⁷³.

To identify AD-related genes/AS events, we fitted a linear regression model with AD diagnosis (categorical outcomes) and surrogate variables calculated by SVA ⁷⁶. AD-related genes/AS events were defined whose coefficients of the ‘diagnosis’ term significantly deviated to zero. To compare the correlation coefficients that evaluated the correlation between AD-associated and age-associated AS changes in males and females independently, we performed Fisher’s z-transformation on Pearson correlation coefficients using R packages *cocor* ⁸³.

To capture the key age-associated splicing changes in males and females that particularly involved in AD pathogenesis, we focused on in the frontal cortex that relevant to AD and further use the sBASEs for AD prediction. We randomly divided 90% samples for training sets and 10% for test sets and repeated for 100 iterations. In each iteration, we performed a 5-fold cross-validation with the feature selection approach called Recursive Feature Elimination (RFE) via R packages *caret* ⁸⁴. Model performance was evaluated by the area under the curve (AUC). Random forest (RF) and support vector machine (SVM) with a linear kernel were used for classification (Fig. 4D and fig. S8D). To evaluate the final selected features that crucially contributed to AD prediction, we calculated the importance of each feature in each iteration (i.e., mean decrease accuracy, MDA) and then ranked the features by the averaged MDAs across 100 iterations. As negative controls, the sex-stratified model predicted AD patients in females and males separately using randomly selected AS events, and the merge-sexes model predicted AD patients in all samples regardless of sex labels using sBASEs.

Transcriptome-wide association analysis with AD pathology

We used a linear regression model to identify the associations between AS events and AD pathology. The pathology includes tangles and plaque stages with continuous outcomes. Age was included as one of the covariates. This association analysis was performed for each sex.

Construction of AS regulatory network

Splicing-related genes were extracted from the GO_RNA_SPLICING term based on MsigDB (c5.all.v7.4.symbols.gmt). Networks between age-associated splicing factors (SFs) and sBASEs in each sex were constructed with the following three steps.

Firstly, we examined the significant correlations between the expression of SFs and splicing of AS events during aging. We conducted spearman’s correlation test between TPMs of SFs and PSIs of AS events during aging using the threshold p-value < 0.05.

Secondly, we chose the AS events significantly perturbed by RBP knockdown for downstream analysis. RBP shRNA RNA-seq bam files of 200 RBPs in HepG2 and K562 were downloaded from ENCODE Consortium. The TPM and PSI values were also quantified by Paean ⁷³. Differential analysis between control and shRNA groups was performed by a linear regression model:

$$PSI_{ij} \sim \mu_i + \beta_G Group_j + \varepsilon_{ij} \quad (8)$$

where $Group_j$ represents a categorical variable labeled by “Control” or “shRNA”. AS events significantly perturbed by RBP knockdown were defined by the estimated coefficient β_G significantly deviated from zero.

Thirdly, we identified the AS events with SF binding motifs around the alternative splice sites (SS). We used DeepBind ⁸⁵ to predict the specific binding scores of RBPs within the upstream to downstream region adjacent to the alternative SS. BEDTools *getfasta* ⁸⁶ extracted sequences from specific genomic regions in a .bed6 file. We first subdivided the sequences into multiple subsequences (window = 40nt; step = 1nt) and calculated the binding scores with the parameter ‘average’ of each subsequence. To identify whether there were binding signals in the upstream to downstream 300nt region adjacent to the alternative splice sites, we rearranged the binding scores

in each region into small bins (width = 20), and tested the median differences between the bins with maximum and minimum averaged binding scores by Wilcoxon rank-sum test. The AS events whose maximum binding scores > 1 and p-value < 0.05 were reserved. Taken together, we required the regulatory connections between the RBPs and AS events that pass all the individual criteria as judged by all three types of experiments with different techniques. The networks between age-associated SFs and sBASEs were constructed and figured by Cytoscape ⁸⁷.

Regulation of hormone receptors to the target genes

To examine the estrogen regulation using the RNA-seq dataset, we downloaded the RNA-seq datasets from MCF-7 cell line stimulated by estrogen receptor agonist (propyl pyrazole triol, PPT) in three biological replicates (GEO ID: GSE203058) ⁴⁸. Additional datasets (ID: GSE89888) are downloaded to confirm our findings in engineered MCF-7 cell lines ⁴⁹ with wildtype and two constitutively active mutants (Y537S, and D538G) of ESR1. These cells were treated with 1nM estradiol (E2) or vehicle control (veh) for 24 hours before collecting the total RNAs and subjected to transcriptome profiling by RNA-seq. We analyzed these data using the same pipelines in earlier Alzheimer's disease datasets (Fig. 4) ⁴. The GE changes under estrogen treatments were evaluated in a linear regression model using the p-values of the coefficients deviated from zero. To confirm the expression changes of splicing factors in other cells or tissues, we downloaded the FPKM table from GSE86609 that analyzed the ER knockout mice with estrogen treatments in mouse brain arcuate nucleus (ARCs) ⁵⁰.

To further identify the regulation of the estrogen and androgen using ChIP-seq dataset (Fig. 5E ⁵ and Fig. 6H ⁶), we evaluated the binding scores of ESR1 and AR to the promoter regions of the target genes (i.e., +/- 1kb of the transcription start sites, TSS). The binding score of each target gene (i.e., $-10 \times \log_{10}[\text{MACS Q-value}]$ from ChIP-seq analysis) was calculated by the ChIP-Atlas ⁸⁸, an integrative database of public ChIP-seq data. The ratios in Fig. 5E ⁵ were calculated by the binding scores of each SFs divided by the median binding score of age-associated SFs common in both sexes. Additionally, the ESR1 ChIP-seq peak files in WT mice treated by estrogen were also downloaded from the Gene Expression Omnibus (ID: GSE3645) ⁵¹. The peaks were annotated using ChIPSeeker ⁸⁹.

Time series and breakpoint analysis

Autoregressive Integrated Moving Average (ARIMA) model is one of the linear models for forecasting chronological trends. For non-stationary data, *auto.arima()* function from R package *forecast* conducts a stepwise procedure to search for the best model using the smallest Akaike Information Criterion (AIC). This procedure is a variation of the Hyndman-Khandakar algorithm ⁵².

We first corrected the TPM/PSI values for each gene/AS event by removing the unknown confounding factors except for the sex and age effect via R package *limma*, and averaged the values of the samples with the same sex and age. Then, we applied ARIMA models to figure out genes with potential chronological trends during aging in each sex. In brief, the input data was first transformed into z-ordered objects (R package *zoo*) and the best model was stepwise optimized by AIC (*auto.arima* function from R package *Forecast*). We filtered out the genes/AS events with the significant trend for further analysis by non-seasonal difference order higher than zero ($D > 0$) and the sum of AR or MA coefficient higher than zero ($AR + MA > 0$).

Furthermore, we calculated the rate of change at each age point and estimated the breakpoints during aging. The rate of change at each age i was evaluated by the differences between bins in $(i - w) \sim i$ and $i \sim (i + w)$ in the following steps. w indicated a series of window spans from 5 to 15 years old. The p-values were first tested by MANOVA of three PCs after reducing dimensionality in the PCA process and then smoothed by LOESS regression with the smoothing bandwidth in the range from 0.25 to 0.75 (with interval 0.05). The p-values of each window span were integrated by

average $-\log_{10}$ transformed p-values and smoothed by LOESS regressions with bandwidth = 0.5. The final rate of changes at each age point was defined by the averaged values through all window spans calculated in the previous step. Moreover, we determined the breakpoints as the age points at the global maximums, as well as the local maximums whose distance to the nearest minimum relative to the global maximum was more than 10%. This model was conducted based on the approach from a previously published study ¹⁶ (URL: <https://github.com/UcarLab/SexDimorphismNatureCommunications>).

Identification of aging-modulated genes (AMGs)

We performed the following steps to identify the AMGs that serve as the key contributors to molecular aging (**Fig. 6D**). We randomly trimmed 20% of chronological genes 200 times and repeated breakpoint analysis in each iteration. Wilcoxon signed-rank test was used to compare the changing rate after removal with that calculated before removal. Using a p-value < 0.05, we obtained the set of genes that could significantly alter the aging rate. The AMGs were defined as the set of genes that were significantly enriched from the 200 random samplings (as judged by the Fisher exact test using a p-value < 0.05). The tissues with more than 10 chronological genes were selected for AMGs identification process. Functional enrichment of sex-biased AMGs was based on MSigDB, and the network between GO terms and genes was constructed by Cytoscape ⁸⁷. GE patterns during aging were shown by the R package *pheatmap*.

Our results observed clusters of AMGs common in both sexes in whole blood tissue with sex-specific changing patterns (i.e., some genes increased first at young ages and then decreased during male aging, whereas these genes monotonically decreased during female aging). To identify AMGs with this expression pattern, we divided the age points into 9 age windows (e.g., 20-25, 26-30, 31-35, etc.) and averaged the ARIMA-fitted expression levels of the samples within each age window. The AMGs increased at young ages were defined by the mean of fitted expression levels at 31-40 larger than at 20-30.

Data and materials availability

GTEX data were obtained through dbGaP (accession number phs000424.v8.p2). The source RNA-seq data (*.fastq) of AD and control samples were obtained from Gene Expression Omnibus (accession number GSE174367). The RNA-seq and ChIP-seq data with estrogen treatment in ER mutant and wild type cells were downloaded from Gene Expression Omnibus (accession number GSE203058, GSE89888, GSE86609, and GSE3645). Pre-mapped RNA-seq data (*.bam) of AD and control samples were obtained from ROSMAP (accession number syn8540863). shRNA-seq data of 200 RBPs in HepG2 and K562 cell lines were downloaded from ENCODE consortium (<https://www.encodeproject.org>). Molecular Signatures Database, (<http://software.broadinstitute.org/gsea/msigdb/collections.jsp>). The source code is available at <https://github.com/rnasys/Sex-dimorphic-aging> and from Zenodo <https://doi.org/10.5281/zenodo.10970995>. All data are available in the main text or the supplementary materials.

Acknowledgements

The authors want to thank Dr. Yue Hu and Jiefu Li in Wang Lab for their discussions and comments. We also thank Prof. Guoqing Zhang in SINH for his assistance with public data access and Ruijie Yao for helping download RBP shRNA-seq bam files from ENCODE Consortium and providing Paean scripts for multi-processing on the CPU-GPU platform.

Additional information

Funding

This work is supported by the Strategic Priority Research Program of CAS (XDB38040100), the Natural Science Foundation of China (91940303, 32030064, and 31730110), and the National Key Research and Development Program of China (2018YFA0107602) to Z.W.. Z.W. is supported by the Starry Night Science Fund at Shanghai Institute for Advanced Study of Zhejiang University (SN-ZJU-SIAS-009) and the type A CAS Pioneer 100-Talent program. This work is also supported by the National Key Research and Development Program of China to X.L. (2021YFA0805200, 2019YFC1315804), and the National Natural Science Foundation of China to X.L. (31970554).

Author contributions

S.W. and Z.W. conceived and designed the study. S.W. analyzed the data and conducted the statistical analysis. S.W. and Z.W. wrote the manuscript. D.D. X.L. provided key suggestions for study design and statistical analysis. D.D. and X.L. downloaded the GTEx RNA-seq and phenotype data.

Conceptualization: SW, ZW

Methodology: SW, DD

Investigation: SW

Visualization: SW

Supervision: XL, ZW

Writing—original draft: SW, ZW

Writing—review & editing: SW, DD, XL, ZW

Additional files

Supplemental Figures [↗](#)

Supplemental Tables [↗](#)

References

- 1 Westergaard D., Moseley P., Sørup F., Baldi P., Brunak S (2019) **Population-wide analysis of differences in disease progression patterns in men and women** *Nature Communications* **10** [Google Scholar](#)
- 2 Khrantsova E. A., Davis L. K., Stranger B. E (2019) **The role of sex in the genomics of human complex traits** *Nature Reviews Genetics* **20**:173–190 [Google Scholar](#)
- 3 Regitz-Zagrosek V (2012) **Sex and gender differences in health: Science & Society Series on Sex and Science** *EMBO reports* **13**:596–603 [Google Scholar](#)
- 4 Ferretti M. T., et al. (2018) **Sex differences in Alzheimer disease—the gateway to precision medicine** *Nature Reviews Neurology* **14**:457–469 [Google Scholar](#)
- 5 Jazin E., Cahill L (2010) **Sex differences in molecular neuroscience: from fruit flies to humans** *Nature Reviews Neuroscience* **11**:9–17 [Google Scholar](#)
- 6 Colafella K. M. M., Denton K. M (2018) **Sex-specific differences in hypertension and associated cardiovascular disease** *Nature Reviews Nephrology* **14**:185–201 [Google Scholar](#)
- 7 Klein S. L., Flanagan K. L (2016) **Sex differences in immune responses** *Nature Reviews Immunology* **16**:626–638 [Google Scholar](#)
- 8 Li C. H., Haider S., Shiah Y.-J., Thai K., Boutros P. C (2018) **Sex differences in cancer driver genes and biomarkers** *Cancer research* **78**:5527–5537 [Google Scholar](#)
- 9 Park M. D (2020) **Sex differences in immune responses in COVID-19** *Nature Reviews Immunology* **20**:461–461 [Google Scholar](#)
- 10 Takahashi T., et al. (2020) **Sex differences in immune responses that underlie COVID-19 disease outcomes** *Nature* **588**:315–320 [Google Scholar](#)
- 11 Sampathkumar N. K., et al. (2020) **Widespread sex dimorphism in aging and age-related diseases** *Human genetics* **139**:333–356 [Google Scholar](#)
- 12 Hägg S., Jylhävä J (2021) **Sex differences in biological aging with a focus on human studies** *eLife* **10**:e63425 <https://doi.org/10.7554/eLife.63425> | [Google Scholar](#)
- 13 Austad S. N., Fischer K. E (2016) **Sex differences in lifespan** *Cell metabolism* **23**:1022–1033 [Google Scholar](#)
- 14 Davis E. J., et al. (2021) **Sex-specific association of the X chromosome with cognitive change and tau pathology in aging and Alzheimer disease** *JAMA neurology* **78**:1249–1254 [Google Scholar](#)
- 15 Ferretti M. T., Santuccione Chadha A (2021) **The missing x factor in alzheimer disease** *Nature Reviews Neurology* **17**:727–728 [Google Scholar](#)

- 16 Márquez E. J., et al. (2020) **Sexual-dimorphism in human immune system aging** *Nature communications* **11**:1–17 [Google Scholar](#)
- 17 Berchtold N. C., et al. (2008) **Gene expression changes in the course of normal brain aging are sexually dimorphic** *Proceedings of the National Academy of Sciences* **105**:15605–15610 [Google Scholar](#)
- 18 Zhou X., et al. (2023) **Molecular differences in brain regional vulnerability to aging between males and females** *Frontiers in Aging Neuroscience* **15** [Google Scholar](#)
- 19 Huang Z., et al. (2021) **Effects of sex and aging on the immune cell landscape as assessed by single-cell transcriptomic analysis** *Proceedings of the National Academy of Sciences* **118** [Google Scholar](#)
- 20 Han Y., et al. (2022) **Proteogenomics reveals sex-biased aging genes and coordinated splicing in cardiac aging** *American Journal of Physiology-Heart and Circulatory Physiology* **323**:H538–H558 [Google Scholar](#)
- 21 GTEx Consortium (2020) **The GTEx Consortium atlas of genetic regulatory effects across human tissues** *Science* **369**:1318–1330 [Google Scholar](#)
- 22 Oliva M., et al. (2020) **The impact of sex on gene expression across human tissues** *Science* **369** [Google Scholar](#)
23. Khodursky S., et al. (2022) **Sex differences in interindividual gene expression variability across human tissue** *PNAS nexus* [Google Scholar](#)
- 24 Zeng L., et al. (2020) **Transcriptome analysis reveals the difference between “ healthy ” and “common” aging and their connection with age-related diseases** *Aging cell* **19**:e13121 [Google Scholar](#)
- 25 Jia K., Cui C., Gao Y., Zhou Y., Cui Q (2018) **An analysis of aging-related genes derived from the genotype-tissue expression project (GTEx)** *Cell death discovery* **4**:1–14 [Google Scholar](#)
- 26 Gallego-Paez L. M., et al. (2017) **Alternative splicing: the pledge, the turn, and the prestige** *Human genetics* **136**:1015–1042 [Google Scholar](#)
- 27 Wang Z., Burge C. B (2008) **Splicing regulation: from a parts list of regulatory elements to an integrated splicing code** *Rna* **14**:802–813 [Google Scholar](#)
- 28 Baralle F. E., Giudice J (2017) **Alternative splicing as a regulator of development and tissue identity** *Nature reviews Molecular cell biology* **18**:437–451 [Google Scholar](#)
- 29 Nikom D., Zheng S (2023) **Alternative splicing in neurodegenerative disease and the promise of RNA therapies** *Nat Rev Neurosci* **24**:457–473 <https://doi.org/10.1038/s41583-023-00717-6> | [Google Scholar](#)
- 30 Wang Y., Bao Y., Zhang S., Wang Z (2020) **Splicing dysregulation in cancer: from mechanistic understanding to a new class of therapeutic targets** *Science China Life Sciences* **63**:469–484 [Google Scholar](#)

- 31 Karlebach G., et al. (2020) **The impact of biological sex on alternative splicing** *BioRxiv* [Google Scholar](#)
- 32 Wang K., et al. (2018) **Comprehensive map of age-associated splicing changes across human tissues and their contributions to age-associated diseases** *Scientific reports* **8**:1–12 [Google Scholar](#)
- 33 Bhadra M., Howell P., Dutta S., Heintz C., Mair W. B (2020) **Alternative splicing in aging and longevity** *Human genetics* **139**:357–369 [Google Scholar](#)
- 34 Angarola B. L., Anczuków O (2021) **Splicing alterations in healthy aging and disease** *Wiley Interdisciplinary Reviews: RNA* [Google Scholar](#)
- 35 Lopes-Ramos C. M., et al. (2020) **Sex differences in gene expression and regulatory networks across 29 human tissues** *Cell reports* **31**:107795 [Google Scholar](#)
- 36 Chan S., Gomes A., Singh R. S (2020) **Is menopause still evolving? Evidence from a longitudinal study of multiethnic populations and its relevance to women’s health** *BMC Women’s Health* **20**:1–15 [Google Scholar](#)
- 37 Ober C., Loisel D. A., Gilad Y (2008) **Sex-specific genetic architecture of human disease** *Nature Reviews Genetics* **9**:911–922 [Google Scholar](#)
- 38 Melé M., et al. (2015) **The human transcriptome across tissues and individuals** *Science* **348**:660–665 <https://doi.org/10.1126/science.aaa0355> | [Google Scholar](#)
- 39 Seidler R. D., et al. (2010) **Motor control and aging: links to age-related brain structural, functional, and biochemical effects** *Neuroscience & Biobehavioral Reviews* **34**:721–733 [Google Scholar](#)
- 40 Balliu B., et al. (2019) **Genetic regulation of gene expression and splicing during a 10-year period of human aging** *Genome biology* **20**:1–16 [Google Scholar](#)
- 41 Ferreira P. G., et al. (2018) **The effects of death and post-mortem cold ischemia on human tissue transcriptomes** *Nature communications* **9** [Google Scholar](#)
- 42 Wucher V., Sodaei R., Amador R., Irimia M., Guigó R (2023) **Day-night and seasonal variation of human gene expression across tissues** *PLoS Biology* **21**:e3001986 [Google Scholar](#)
- 43 Li Y., Ge X., Peng F., Li W., Li J. J (2022) **Exaggerated false positives by popular differential expression methods when analyzing human population samples** *Genome biology* **23**:1–13 [Google Scholar](#)
- 44 Xiong J., et al. (2022) **FSH blockade improves cognition in mice with Alzheimer’s disease** *Nature* :1–7 [Google Scholar](#)
- 45 Matera A. G., Wang Z (2014) **A day in the life of the spliceosome** *Nature reviews Molecular cell biology* **15**:108–121 [Google Scholar](#)
- 46 Bergles D. E., Richardson W. D. (2016) **Oligodendrocyte development and plasticity** *Cold Spring Harbor perspectives in biology* **8**:a020453 [Google Scholar](#)
- 47 Horstman A. M., Dillon E. L., Urban R. J., Sheffield-Moore M (2012) **The Role of Androgens and Estrogens on Healthy Aging and Longevity** *The Journals of Gerontology: Series A* **67**:1140–

1152 <https://doi.org/10.1093/gerona/gls068> | [Google Scholar](#)

- 48 Cao L.-B., et al. (2023) **Estrogen receptor α -mediated signaling inhibits type I interferon response to promote breast cancer** *Journal of Molecular Cell Biology* :mjad047 [Google Scholar](#)
- 49 Bahreini A., et al. (2017) **Mutation site and context dependent effects of ESR1 mutation in genome-edited breast cancer cell models** *Breast Cancer Research* **19**:1–10 [Google Scholar](#)
- 50 Yang J. A., Stires H., Belden W. J., Roepke T. A (2017) **The Arcuate Estrogen-Regulated Transcriptome: Estrogen Response Element-Dependent and-Independent Signaling of ER α in Female Mice** *Endocrinology* **158**:612–626 [Google Scholar](#)
- 51 Hewitt S. C., et al. (2014) **Novel DNA motif binding activity observed in vivo with an estrogen receptor α mutant mouse** *Molecular Endocrinology* **28**:899–911 [Google Scholar](#)
- 52 Brockwell P. J., Brockwell P. J., Davis R. A., Davis R. A (2016) **Introduction to time series and forecasting** Springer [Google Scholar](#)
- 53 Durieux J., Wolff S., Dillin A (2011) **The cell-non-autonomous nature of electron transport chain-mediated longevity** *Cell* **144**:79–91 [Google Scholar](#)
- 54 Djebali S., et al. (2012) **Landscape of transcription in human cells** *Nature* **489** [Google Scholar](#)
- 55 Wang E. T., et al. (2008) **Alternative isoform regulation in human tissue transcriptomes** *Nature* **456**:470–476 [Google Scholar](#)
- 56 Roberts R., et al. (2012) **The incidence of MCI differs by subtype and is higher in men: the Mayo Clinic Study of Aging** *Neurology* **78**:342–351 [Google Scholar](#)
- 57 Foo H., et al. (2021) **Age-and sex-related topological organization of human brain functional networks and their relationship to cognition** *Frontiers in aging neuroscience* **13**:758817 [Google Scholar](#)
- 58 Carithers L. J., et al. (2015) **A novel approach to high-quality postmortem tissue procurement: the GTEx project** *Biopreservation and biobanking* **13**:311–319 [Google Scholar](#)
- 59 Nygaard V., Rødland E. A., Hovig E (2016) **Methods that remove batch effects while retaining group differences may lead to exaggerated confidence in downstream analyses** *Biostatistics* **17**:29–39 [Google Scholar](#)
- 60 Van Nostrand E. L., et al. (2020) **A large-scale binding and functional map of human RNA-binding proteins** *Nature* **583**:711–719 [Google Scholar](#)
- 61 Shen X., et al. (2024) **Nonlinear dynamics of multi-omics profiles during human aging** *Nature aging* :1–16 [Google Scholar](#)
- 62 Ramanujan S. A., Cravens E. N., Krishfield S. M., Kyttaris V. C., Moulton V. R (2021) **Estrogen-Induced hsa-miR-10b-5p Is Elevated in T Cells From Patients With Systemic Lupus Erythematosus and Down-Regulates Serine/Arginine-Rich Splicing Factor 1** *Arthritis & Rheumatology* **73**:2052–2058 [Google Scholar](#)

- 63 Yoshida M., Apte R. S (2022) **NAD⁺ boosting brings tears to aging eyes** *Nature Aging* :1–3 [Google Scholar](#)
- 64 Xiong X., et al. (2021) **Genetic drivers of m6A methylation in human brain, lung, heart and muscle** *Nature Genetics* **53**:1156–1165 [Google Scholar](#)
- 65 Jylhävä J., Pedersen N. L., Hägg S (2017) **Biological age predictors** *EBioMedicine* **21**:29–36 [Google Scholar](#)
- 66 Horvath S., Raj K (2018) **DNA methylation-based biomarkers and the epigenetic clock theory of ageing** *Nature Reviews Genetics* **19**:371–384 [Google Scholar](#)
- 67 Galkin F., Mamoshina P., Kochetov K., Sidorenko D., Zhavoronkov A (2021) **DeepMAge: A methylation aging clock developed with deep learning** *Aging and disease* **12**:1252 [Google Scholar](#)
- 68 Meyer D. H., Schumacher B (2021) **BiT age: A transcriptome-based aging clock near the theoretical limit of accuracy** *Aging Cell* [Google Scholar](#)
- 69 Sayed N., et al. (2021) **An inflammatory aging clock (iAge) based on deep learning tracks multimorbidity, immunosenescence, frailty and cardiovascular aging** *Nature Aging* **1**:598–615 [Google Scholar](#)
- 70 Galkin F., et al. (2020) **Human gut microbiome aging clock based on taxonomic profiling and deep learning** *IScience* **23** [Google Scholar](#)
- 71 Li J., et al. (2023) **Determining a multimodal aging clock in a cohort of Chinese women** *Med* <https://doi.org/10.1016/j.medj.2023.06.010> | [Google Scholar](#)
- 72 Hannum G., et al. (2013) **Genome-wide methylation profiles reveal quantitative views of human aging rates** *Molecular cell* **49**:359–367 [Google Scholar](#)
- 73 Li J., et al. (2018) **Paeon: A parallel transcriptome quantification tool combining gene expression and alternative splicing events using GPU** In: *IEEE International Conference on Bioinformatics and Biomedicine (BIBM)* IEEE pp. 676–679 [Google Scholar](#)
- 74 Song M., et al. (2020) **A review of integrative imputation for multi-omics datasets** *Frontiers in genetics* **11**:570255 [Google Scholar](#)
- 75 Zhang Y., et al. (2020) **Regional variation of splicing QTLs in human brain** *The American Journal of Human Genetics* **107**:196–210 [Google Scholar](#)
- 76 Leek J. T., Johnson W. E., Parker H. S., Jaffe A. E., Storey J. D (2012) **The sva package for removing batch effects and other unwanted variation in high-throughput experiments** *Bioinformatics* **28**:882–883 [Google Scholar](#)
- 77 Dolgalev I. (2020) **msigdb: MSigDB gene sets for multiple organisms in a tidy data format** *R package v.7.0.0* [Google Scholar](#)
- 78 Yu G., Wang L.-G., Han Y., He Q.-Y (2012) **clusterProfiler: an R package for comparing biological themes among gene clusters** *Omics: a journal of integrative biology* **16**:284–287 [Google Scholar](#)

- 79 Yu G., Wang L.-G., Yan G.-R., He Q.-Y (2015) **DOSE: an R/Bioconductor package for disease ontology semantic and enrichment analysis** *Bioinformatics* **31**:608–609 [Google Scholar](#)
- 80 Andrews S. (2010) **Babraham Bioinformatics** Cambridge, United Kingdom: Babraham Institute [Google Scholar](#)
- 81 Langmead B., Salzberg S. L (2012) **Fast gapped-read alignment with Bowtie 2** *Nature methods* **9**:357–359 [Google Scholar](#)
- 82 Dobin A., et al. (2013) **STAR: ultrafast universal RNA-seq aligner** *Bioinformatics* **29**:15–21 [Google Scholar](#)
- 83 Diedenhofen B., Musch J. (2015) **cocor: A comprehensive solution for the statistical comparison of correlations** *PloS one* **10**:e0121945 [Google Scholar](#)
- 84 Kuhn M., Johnson K (2013) **Applied predictive modeling** Springer [Google Scholar](#)
- 85 Alipanahi B., Delong A., Weirauch M. T., Frey B. J (2015) **Predicting the sequence specificities of DNA-and RNA-binding proteins by deep learning** *Nature biotechnology* **33**:831–838 [Google Scholar](#)
- 86 Quinlan A. R., Hall I. M (2010) **BEDTools: a flexible suite of utilities for comparing genomic features** *Bioinformatics* **26**:841–842 [Google Scholar](#)
- 87 Shannon P., et al. (2003) **Cytoscape: a software environment for integrated models of biomolecular interaction networks** *Genome research* **13**:2498–2504 [Google Scholar](#)
- 88 Oki S., et al. (2018) **Ch IP-Atlas: a data-mining suite powered by full integration of public Ch IP-seq data** *EMBO reports* **19**:e46255 [Google Scholar](#)
- 89 Yu G., Wang L.-G., He Q.-Y (2015) **ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization** *Bioinformatics* **31**:2382–2383 [Google Scholar](#)

Author information

Siqi Wang

Shanghai Institute of Nutrition and Health, University of Chinese Academy of Sciences, Chinese Academy of Sciences, Shanghai, China, Department of Integrative Biology and Physiology, University of California, Los Angeles, Los Angeles, United States
ORCID iD: [0000-0002-1491-2962](https://orcid.org/0000-0002-1491-2962)

For correspondence: siqiwang@ucla.edu

Danyue Dong

Shanghai Institute of Nutrition and Health, University of Chinese Academy of Sciences, Chinese Academy of Sciences, Shanghai, China

Xin Li

Shanghai Institute of Nutrition and Health, University of Chinese Academy of Sciences, Chinese Academy of Sciences, Shanghai, China

Zefeng Wang

Shanghai Institute of Nutrition and Health, University of Chinese Academy of Sciences, Chinese Academy of Sciences, Shanghai, China, School of Life Science, Southern University of Science and Technology, Shenzhen, China

For correspondence: wangzf@sustec.edu.cn

Editors

Reviewing Editor

Aditi Gurkar

University of Pittsburgh, Pittsburgh, United States of America

Senior Editor

Pankaj Kapahi

Buck Institute for Research on Aging, Novato, United States of America

Reviewer #2 (Public review):

Summary:

In this manuscript, Wang et al analyze ~17,000 transcriptomes from 35 human tissues from the GTEx database and address transcriptomic variations due to age and sex. They identified both gene expression changes as well as alternative splicing events that differ among sexes. Using breakpoint analysis, the authors find sex dimorphic shifts begin with declining sex hormone levels with males being affected more than females. This is an important pan-tissue transcriptomic study exploring age and sex-dependent changes although not the first one.

Strengths:

- (1) The authors use sophisticated modeling and statistics for differential, correlational and predictive analysis.
- (2) The authors consider important variables such as genetic background, ethnicity, sampling bias, sample sizes, detected genes etc.
- (3) This is likely the first study to evaluate alternative splicing changes with age and sex at a pan-tissue scale.
- (4) Sex dimorphism with age is an important topic and is thoroughly analyzed in this study.

<https://doi.org/10.7554/eLife.102449.2.sa2>

Reviewer #3 (Public review):

Summary:

In this study, Wang et al utilized the available GTEx data to compile a comprehensive analysis that attempt to reveal aging-related sex-dimorphic gene expression as well as alternative splicing changes in human. The key conclusions based upon their analysis are that 1) extensive sex-dimorphisms during aging with distinct patterns of change in gene expression and alternative splicing (AS), and 2) the male-biased age-associated AS events have a stronger association with Alzheimer's disease, and 3) the females-biased events are often regulated by several sex-biased splicing factors that may be controlled by estrogen receptors. They further

performed break-point analysis and reveal in males there are two main breakpoints around ages 35 and 50, while in female only one breakpoint at 45.

Strengths:

This study sets an ambitious goal, leveraging the extensive GTEx dataset to investigate aging-related, sex-dimorphic gene expression and alternative splicing changes in humans. The research addresses a significant question, as our understanding of sex-dimorphic gene expression in the context of human aging is still in its early stages. Advancing our knowledge of these molecular changes is vital for identifying therapeutic targets for age-related diseases and extending human healthspan. The study is highly comprehensive, and the authors are commendable for their attempted thorough analysis of both gene expression and alternative splicing-an area often overlooked in similar studies.

<https://doi.org/10.7554/eLife.102449.2.sa1>

Author response:

The following is the authors' response to the original reviews

Public Reviews:

Reviewer #1 (Public review):

Summary:

Wang et al. investigate sexual dimorphic changes in the transcriptome of aged humans. This study relies upon analysis of the Genotype-Tissue Expression dataset that includes 54 tissues from human donors. The authors investigate 17,000 transcriptomes from 35 tissues to investigate the effect of age and sex on transcriptomic variation, including the analysis of alternative splicing. Alternative splicing is becoming more appreciated as an influence in the aging process, but how it is affected by sexual dimorphism is still largely unclear. The authors investigated multiple tissues but ended up distilling brain tissue down to four separate regions: decision, hormone, memory, and movement. Building upon prior work, the authors used an analysis method called principal component-based signal-to-variation ratio (pcSVR) to quantify differences between sex or age by considering data dispersion. This method also considers differentially expressed genes and alternative splicing events.

Strengths:

- (1) The authors investigate sexual dimorphism on gene expression and alternative splicing events with age in multiple tissues from a large publicly available data set that allows for reanalysis.*
- (2) Furthermore, the authors take into account the ethnic background of donors. Identification of agingmodulating genes could be useful for the reanalysis of prior data sets.*

Weaknesses:

The models built off of the GTEx dataset should be tested in another data set (ex. Alzheimer's disease) where there are functional changes that can be correlated. Gene-length-dependent transcription decline, which occurs with age and disease, should also be investigated in this data set for potential sexual dimorphism.

We appreciate the reviewer's constructive feedback and acknowledgment of the strengths of our study. The detailed results are included in the 'Recommendations for the authors' from the editorial office. Below we summarize our feedback that address the concerns of this reviewer:

(1) Independent Alzheimer's disease (AD) datasets:

We acknowledge the importance of validating our models beyond GTEx to assess their generalizability aging to Alzheimer's disease. While GTEx provides valuable transcriptomic data across multiple tissues, it lacks direct functional assessments linked to disease states. We have already analyzed RNA-seq data from ROSMAP and GEO in Figure 4, focusing on sex-biased gene expression and splicing changes between aging and AD. The results showed a male-biased association with Alzheimer's disease at AS resolution, indicating that the AS changes during aging could contribute more to AD in males than females. We added a highlight to this analysis in the manuscript (Pages 6-7).

(2) Sexual dimorphism in Gene-Length-Dependent Transcription Decline (GLTD)

We appreciate the reviewer's suggestion to explore gene-length-dependent transcription decline (GLTD), which has been implicated in both aging and disease. As the reviewer suggested, our analysis revealed that GLTD exhibits sex-biased patterns in different tissues, aligning with recent literature on sex-dimorphic transcriptional aging. Our findings also revealed that longer genes with greater transcriptional decline are enriched in AD-related pathways. We have incorporated this new analysis in the 'Recommendations for the authors' in Author response image 5-6 and expanded the discussion of the biological relevance.

Reviewer #2 (Public review):

Summary:

In this manuscript, Wang et al analyze ~17,000 transcriptomes from 35 human tissues from the GTEx database and address transcriptomic variations due to age and sex. They identified both gene expression changes as well as alternative splicing events that differ among sexes. Using breakpoint analysis, the authors find sex dimorphic shifts begin with declining sex hormone levels with males being affected more than females. This is an important pan-tissue transcriptomic study exploring age and sex-dependent changes although not the first one.

Strengths:

- (1) The authors use sophisticated modeling and statistics for differential, correlational, and predictive analysis.*
- (2) The authors consider important variables such as genetic background, ethnicity, sampling bias, sample sizes, detected genes, etc.*
- (3) This is likely the first study to evaluate alternative splicing changes with age and sex at a pan-tissue scale.*
- (4) Sex dimorphism with age is an important topic and is thoroughly analyzed in this study.*

Weaknesses:

- (1) The findings have not been independently validated in a separate cohort or through experiments. Only selective splicing factor regulation has been verified in other studies.*
- (2) It seems the authors have not considered PMI or manner of death as a variable in their analysis.*

(3) The manuscript is very dense and sometimes difficult to follow due to many different types of analyses and correlations.

(4) Short-read data can detect and quantify alternative splicing events with only moderate confidence and therefore the generalizability of these findings remains to be experimentally validated.

We appreciate the thorough review and thoughtful feedback. We have addressed the reviewer's concerns and added clarification. The detailed results are included in Recommendations for the authors. Here are the summaries.

(1) Challenge of independent validation in separate cohorts

- The GTEx dataset includes the most comprehensive transcriptome resource for studying population-level differences in age and sex across tissues, particularly including large-scale brain samples. This provides a unique opportunity to analyze sex-dimorphic aging and the relevance of age-associated diseases. Several technical issues, including cell type heterogeneity, postmortem artifacts, as well as sequencing biases, lead to technical challenges in different cohorts.
- As the reviewer mentioned, we analyzed transcriptomic data from Shen et al. (2024) and compared them with GTEx results (Author response image 2). Limited overlap in differentially expressed genes again highlighted the challenges in cross-dataset validation due to the differences in cell composition and data processing (peripheral blood mononuclear cells (PBMCs) vs whole blood).
- Due to the limited human brain transcriptome data covering different age and sex groups, we found mouse hippocampus datasets from Mass spectrometry (MS), including young and old, as well as female and male groups. The results validated the expression of splicing factors in brain (Author response image 9). This cross-species consistency supports the robustness of our findings in human brain aging.

(2) Effects of Postmortem Interval, Manner of Death, and Time of Death

- We agree that the sample collections could introduce confounding effects. To address this, we calculated the correlations between the confounding factors with Postmortem Interval (PMI), Manner of Death (DTHMNNR), or Time of Death (DTHTIME and DTHSEASON). We observed strong correlations in some surrogate variables in most tissues, indicating that those factors could be well-regressed during our analysis (Recommendations for the authors, Figure S4 and R8).
- In addition, we re-evaluated our analyses while incorporating PMI as a covariate in our models. Our results align with our initial findings (Author response image 1), suggesting that age- and sex-dependent transcriptomic changes are not strongly confounded by PMI and confirming that our model has controlled PMI. These results are detailed in 'Recommendations for the authors' and included in Figure S4C-E with the description in text, Page 5.

(3) Readability of manuscript and flow of analyses

- In summary, our study first examined global alternative splicing (AS) and gene expression (GE) across all tissues before focusing on specific regions for deeper insights. To improve clarity, we have made the following revisions:
- Add clearer statements when transitioning between all-tissue and brain-specific analyses (Page 6-7).

- Modify the subtitle of Results to highlight all-tissue vs. brain analyses (Page 6).
- These refinements could enhance the manuscript's structure, making the flow of analysis and conclusions more intuitive for readers.

(4) Limitations of short-read RNA-seq for splicing analysis

- Short-read RNA-seq provides only moderate confidence in detecting and quantifying full-length isoforms. However, its higher sequencing depth makes it more suitable for quantifying changes in alternative splicing (AS) events.
- Our analysis focused on splicing event-level quantification, applying stringent filters and using our GPU-based tool, which showed strong concordance with RT-PCR and other pipelines. Therefore, we also cited and included the updated Paeon manuscript that benchmarks its performance in AS analysis.

Reviewer #3 (Public review):

Summary:

In this study, Wang et al utilized the available GTEx data to compile a comprehensive analysis that attempt to reveal aging-related sex-dimorphic gene expression as well as alternative splicing changes in humans.

The key conclusions based on their analysis are that.

(1) extensive sex-dimorphisms during aging with distinct patterns of change in gene expression and alternative splicing (AS), and

(2) the male-biased age-associated AS events have a stronger association with Alzheimer's disease, and (3) the female-biased events are often regulated by several sex-biased splicing factors that may be controlled by estrogen receptors. They further performed break-point analysis and revealed that in males there are two main breakpoints around ages 35 and 50, while in females, there is only one breakpoint at 45.

Strengths:

This study sets an ambitious goal, leveraging the extensive GTEx dataset to investigate aging-related, sexdimorphic gene expression and alternative splicing changes in humans. The research addresses a significant question, as our understanding of sex-dimorphic gene expression in the context of human aging is still in its early stages. Advancing our knowledge of these molecular changes is vital for identifying therapeutic targets for age-related diseases and extending the human health span. The study is highly comprehensive, and the authors are commendable for their attempted thorough analysis of both gene expression and alternative splicing - an area often overlooked in similar studies.

We thank this reviewer for the insightful review and recognition of our study's significance. We agree with the reviewer on how to examine sex-dimorphic gene expression and alternative splicing in aging by using the GTEx dataset. This is indeed an essential aspect of developing potential therapeutic targets for agerelated diseases to promote human health span.

Weaknesses:

Due to the inherent noise within the GTEx dataset - which includes numerous variables beyond aging and sex - there are significant technical concerns surrounding this study.

Additionally, the lack of crossvalidation with independent, existing data raises questions about whether the observed gene expression changes genuinely reflect those associated with human aging. For instance, the break-point analysis in this study identifies two major breakpoints in males around ages 35 and 50, and one breakpoint in females at age 45; however, these findings contradict a recent multi-omics longitudinal study involving 108 participants aged 25 to 75 years, where breakpoint at 44 and 60 years was observed in both male and females (Shen et al, 2024). These issues cast doubt on the robustness of the study's conclusions. Specific concerns are outlined below:

References:

Ferreira PG, Muñoz-Aguirre M, Reverter F, Sá Godinho CP, Sousa A, Amadoz A, Sodaei R, Hidalgo MR, Pervouchine D, Carbonell-Caballero J et al (2018) The effects of death and post-mortem cold ischemia on human tissue transcriptomes. *Nature Communications* 9: 490.

Shen X, Wang C, Zhou X, Zhou W, Hornburg D, Wu S, Snyder MP (2024) Nonlinear dynamics of multiomics profiles during human aging. *Nature Aging*.

Wucher V, Sodaei R, Amador R, Irimia M, Guigó R (2023) Day-night and seasonal variation of human gene expression across tissues. *PLOS Biology* 21: e3001986.

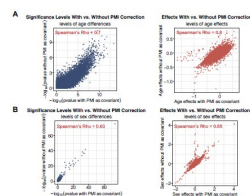
(1) The primary method used in this study is linear regression, incorporating age, sex, and age-by-sex interactions as covariates, alongside other confounding factors (such as ethnicity) as unknown variables. However, the analysis overlooks two critical known variables in the GTEx dataset: time of death (TOD) and postmortem interval (PMI). Both TOD and PMI are recorded for each sample and account for substantial variance in gene expression profiles. A recent study by Wucher et al. (Wucher et al, 2023) demonstrated the powerful impact of TOD on gene expression by using it to reconstruct human circadian and even circannual datasets. Similarly, Ferreira et al. (Ferreira et al, 2018) highlighted PMI's influence on gene expression patterns. Without properly adjusting for these two variables, confidence in the study's conclusions remains limited at best.

We appreciate the reviewer for raising this important point regarding the impact of post-mortem interval (PMI) and time of death (TOD) on gene expression, including the death seasons (DTHSEASON) and daytime (DTHTIME). To address this point, we carefully evaluated whether our linear model controlled for these factors as potential confounders.

Our results showed that PMI and TOD significantly correlated with the estimated covariates in most tissues, suggesting that their effects could be effectively regressed out using our model (Figure S4). As the reviewers and editors suggested, we have now included this correlation analysis in the updated Figure S4C-E and the text in the Results section, citing relevant literature [1,2] (Page 5).

Author response image 1.

The results of differential gene expression analysis with vs without the inclusion of PMI correction as a known covariate. The scatter plots show the correlations of significance levels (pvalues, left panel) and effect sizes (coefficients, right panel) of sex (A) and age (B). Whole-blood tissue is used as an example.



In addition, we did the differential analysis that incorporated PMI as a covariate in the regression models and re-evaluated the age- and sex-related transcriptomic changes. Using WholeBlood gene expression as an example, our revised analysis shows that the inclusion of PMI in the covariates has minimal impact on the significance levels and effects of sex and age (i.e., p-values and coefficients, respectively), indicating that our findings are robust using confounding factors (Author response image 1).

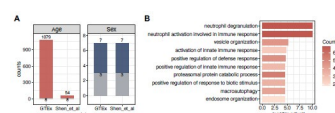
(2) To demonstrate that their analysis is robust and that the covariates TOD and PMI are otherwise negligible - the authors should cross-validate their findings with independent datasets to confirm that the identified gene expression changes are reproducible for some tissues. For instance, the recent study by Shen et al. (Shen et al., 2024) in Nature Aging offers an excellent dataset for cross-validation, particularly for blood samples. Comparing the GTEx-derived results with this longitudinal transcriptome dataset would enable verification of gene expression changes at both the individual gene and pathway levels. Without such validation, confidence in the study's conclusions remains limited.

We thank the reviewer for the insightful suggestion regarding cross-validation with independent datasets. We understand that validating findings across datasets is crucial for ensuring robustness. As the reviewers suggested, we see whether there are some shared findings in the GTEx data with the study by Shen et al. (2024) in Nature Aging. However, after performing comparisons with our GTEx results in whole blood tissue, we found that the overlaps of differentially expressed genes are limited (Fig. 3). In our results, we found a large proportion of age-associated genes in the GTEx data, whereas just 54 genes are age-associated from Shen et al.'s PBMC data. 3 in 7 genes are differentially expressed in both datasets (Fig. 3A). Additionally, we performed the functional enrichment analysis on the GTEx-specific age-associated genes.

We observed a strong enrichment in the biological pathways related to neutrophil functions and innate immune responses, which are specific to the cell compositions in whole blood rather than PBMC (Fig. 3B).

Author response image 2.

The comparison between the gene expression of whole blood tissue from GTEx and PBMCs from Shen et al. (A) The bar plot shows the number of age (left panel) or sex-associated (right panel) genes in the two datasets. The grey bars highlight the proportion of overlapped genes in both datasets. (B) The top 10 significantly enriched biological processes in the GTEx-specific age-associated genes. The color bar shows the number of age-associated genes in specific pathways.



These discrepancies highlighted the crucial factors in cross-dataset comparison:

- Cell compositions: GTEx used whole blood, which contains all blood components, including neutrophils and erythrocytes, whereas PBMCs contain lymphocytes and monocytes. Under the influence of granulocytes and red blood cells in whole blood, the gene expression profiles between these two datasets are different.
- Biological functions: Whole blood includes both innate and adaptive immune components; thus, aging-related gene expression changes in whole blood may include a broader systemic response than those in PBMCs. This difference in biological context contributes to the observed variation in the differentially expressed genes, as demonstrated by our functional enrichment analysis (Fig. 3B).
- Sequencing biases and data processing: The two datasets were generated using different RNAseq processing pipelines, including distinct normalization, batch correction, and quantification methodologies. These technical differences may introduce systematic variations that complicate direct cross-validation.

Due to these fundamental problems, a direct one-to-one validation between the two datasets is challenging. We understand the importance of independent dataset validation and appreciate the reviewer's suggestion. However, future studies could be performed more precisely if comparable whole-blood-based datasets are available. In addition, GTEx data provides nearly thousands of samples in whole blood, which is a largescale, comprehensive, and clinically relevant dataset for studying aging-related changes, particularly in innate immunity and inflammation, which are not well captured in PBMCs.

(3) As a demonstration of the lack of such validation, in the Shen et al. study (Shen et al., 2024), breakpoints at 44 and 60 years were observed in both males and females, while this study identifies two major breakpoints in males around ages 35 and 50, and one breakpoint in females at age 45. What caused this discrepancy?

We thank the reviewer and the editors for both coming up with the non-linear multi-omic aging patterns observed by Shen et al. They observed two prominent crests around the ages of 45 and 60 from omics data.

Similarly, we also identified two breakpoints in our analysis, with some differences in specific age breakpoints. These could be the result of sample preparation methods and breakpoint definition. These responses are also included in the editor's recommendations.

Definition of breakpoints vs crests:

- Crests represent age-related molecular changes at each time point across the human lifespan. They indicate the number of molecules that are differentially expressed during aging ($q < 0.05$), without considering individual expression levels.
- Our breakpoints, in contrast, are identified after filtering the chronological trends using the Autoregressive Integrated Moving Average (ARIMA) model. We calculated the rate of change at each age point using the smooth approach and sliding windows. Breakpoints are defined as local maxima where the distance to the nearest minimum, relative to the global maximum. We indeed found some local wide peaks around 60 in some tissues, shown in Figure S10, however, we excluded these due to our strict cutoffs to remove noise.

Differences and similarities between sequenced tissues:

- Whole-blood vs PBMC: In the GTEx RNA-seq data used in our study, whole blood samples from donors were sequenced, whereas their study used PBMCs. Whole blood contains all blood components, including red blood cells, platelets, granulocytes (e.g., neutrophils), lymphocytes, and monocytes, while PBMCs represent a subset of white blood cells, primarily

consisting of lymphocytes (T cells, B cells, NK cells) and monocytes, excluding granulocytes and erythrocytes. As we mentioned in the previous responses, the gene expression changes observed in whole blood capture the contributions of neutrophils and other granulocytes, which are neglected in the PBMC profile (also shown in Figure S11C).

- For the shared tissues in two studies – skin, we looked at the non-linear changes during aging and found the same two breakpoints: 43 and 58.

Novelties in our study:

- Whole blood can serve as a readily accessible resource for testing age-related disease biomarkers without cell separation, making it more practical for clinical applications.
- Our analysis was performed on females and males, respectively. The main object of our analysis is to compare the differences in aging rates between sexes. Our results reveal clear sex-specific differences across multiple human tissues. Therefore, the identified breakpoints may differ when sex effects are not taken into account, highlighting the specificity of our analysis.
- Additionally, our breakpoints are integrated across multiple tissues. Our results showed that there is a large diversity of aging patterns in different tissues.

As the reviewers and editors suggested, we have added the following statements to clarify this distinction in the Discussion section: ‘Our analysis observed the non-linear aging patterns with two breakpoints, which is consistent with recent findings, with differences in specific age points due to sex differences as well as tissue diversities 3.’ (Page 14), and ‘These breakpoints could represent key junctures in the aging process that align with the non-linear patterns of aging and disease progression.’ (Page 15)

(4) Although the alternative splicing analysis is intriguing, the authors did not differentiate between splicing events that alter the protein-coding sequence and those that do not. Many splicing changes occurring in the 5' UTR and 3' UTR regions do not impact protein coding, so it is essential to filter these out and focus specifically on alternative splicing events that can modify protein-coding sequences.

The reviewer raises an important point. In our study, we included the AS events in protein-coding genes to gain a comprehensive understanding of sex-biased age-associated splicing. As the reviewer suggested, focusing on coding-sequence-altering events is particularly relevant to protein function. To address this, we performed an additional analysis to specifically annotate sBASEs occurring within the coding sequence (defined as CDS-altering sBASEs) and reanalyzed their functional pathways and AD-associations (Author response image 3).

Our analysis revealed that most of the sBASEs are relevant to protein-coding sequences (CDS) across multiple tissues (Author response image 3A). We then confirmed our findings using CDS-altering sBASEs. We found that those sBASEs in brain regions were significantly enriched in pathways related to amyloid-beta formation and actin filament organization (Author response image 3B). Notably, male-biased sBASEs in decision-related brain regions were particularly associated with dendrite development and regulation of cell morphogenesis, highlighting the sex-specific roles of sBASEs in brain functions. Additionally, we performed a random forest classification using only CDS-altering sBASEs in AD datasets (Author response image 3C-D), again confirming the malebiased association between aging and AD.

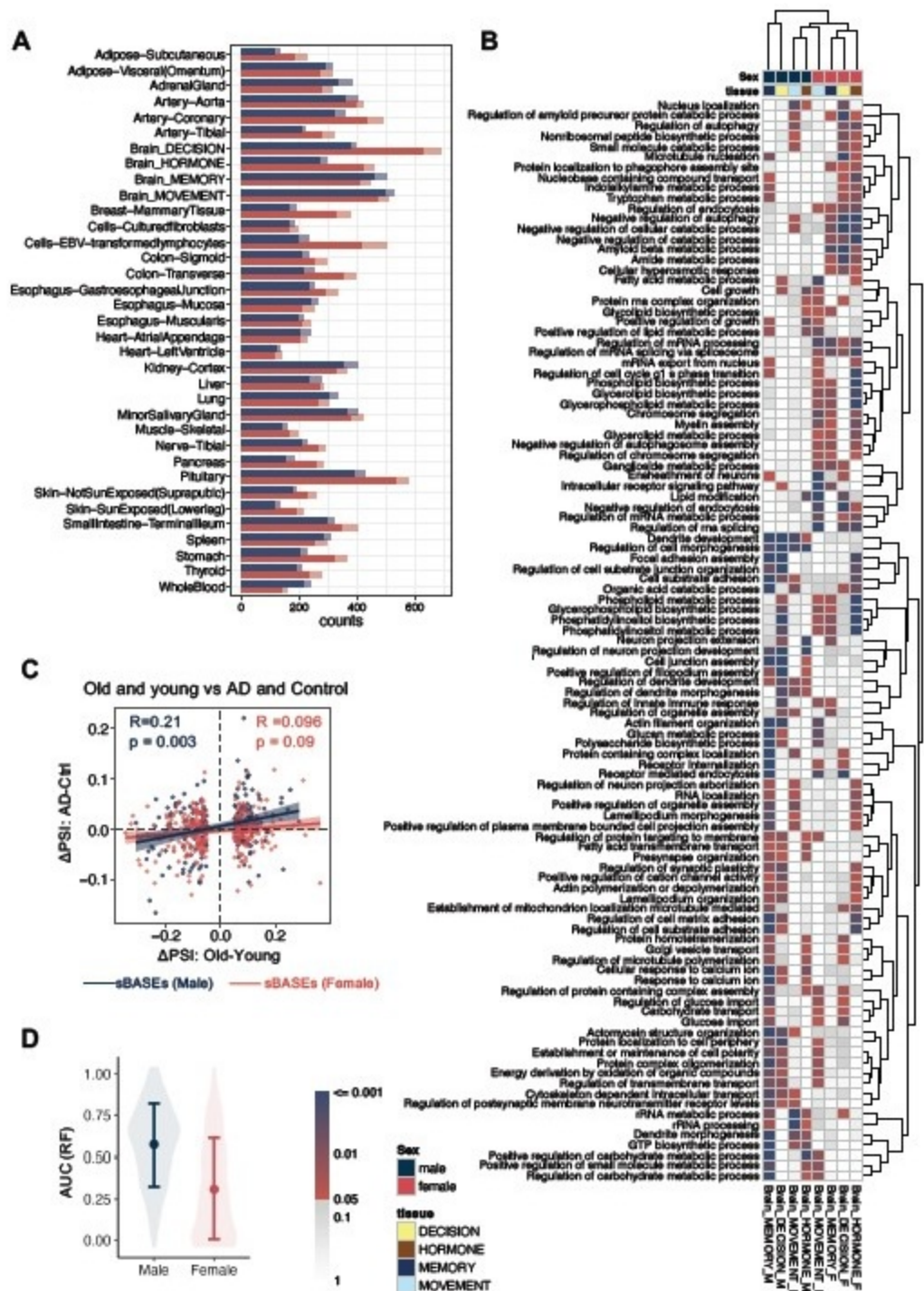
Overall, we found that most of the identified sBASEs could modify protein-coding sequences, and our main conclusions remain consistent even after filtering out non-coding events.

Nevertheless, in addition to AS events that impact protein sequences, alternative splicing in untranslated regions (UTRs) also plays a critical regulatory role. Splicing events in the 5' UTR

can influence translation efficiency by modifying upstream open reading frames (uORFs) or RNA secondary structures, while splicing in the 3'UTR can affect mRNA stability, localization, and translation by altering microRNA binding sites and RNA-binding protein interactions. Given these functional implications, we believe that UTR-targeted AS events should also be considered to supplement the understanding of post-transcriptional gene regulation in future research.

Author response image 3.

The distribution and functional relevance of sBASEs with coding effects. (A) The number of sBASEs and CDS-altering sBASEs across multiple tissues. The deeper bars show the number of sBASEs whose alternative splice sites are located at protein-coding regions. (B) GO biological pathways in each sex and brain region. Heatmap shows the sex-specific pathways that are significantly enriched by CDS-altering sBASEs in more than 2 brain regions and sex. (C) Correlation between AD-associated and age-associated AS changes across the CDS-altering sBASEs that alter protein-coding sequences in females and males. (D) Performances of sex-stratified models predicted by CDS-altering sBASEs in 100 iterations using the random forest approach



(5) One of the study's main conclusions - that "male-biased age-associated AS events have a stronger association with Alzheimer's disease" - is not supported by the data presented in Figure 4A, which shows an association with "regulation of amyloid precursor formation" only in female, not male, alternative splicing genes. Additionally, the gene ontology term "Alzheimer's disease" is absent from the unbiased GO analysis in Figure S6. These discrepancies suggest that the focus on Alzheimer's disease may reflect selective data interpretation rather than results driven by an unbiased analysis.

We thank the reviewer for this point. In our functional analysis, we identified distinct biological processes enriched in female- and male-biased AS genes, such as the regulation of amyloid precursor formation in females and structural constituents of the cytoskeleton in males. However, Alzheimer's disease (AD) is a complex neurodegenerative disorder with multiple pathological mechanisms beyond amyloid-beta (A β) formation, many of which are strongly age-related in both sexes. This complexity motivates us to explore novel relationships between splicing and AD in distinct sexes.

Although Figure 4A shows the enrichment of "regulation of amyloid precursor formation" in female-biased AS events, this does not contradict the broader enrichment of AD-related processes in male-biased AS events. Our disease ontology analysis supports this finding, as male-biased age-associated AS events are enriched in neurodegenerative diseases, including cognitive disorders. Additionally, we considered not only individual GO terms but also the disease-associated transcriptomic signatures from AD-related datasets, which collectively indicate a stronger association in males.

Regarding Figure S6 mentioned by the reviewer, the GO term "Alzheimer's disease" is not explicitly listed in the heatmap because we filtered the pathways that are consistently enriched in multiple tissues. As noted in the figure legend, we only displayed sex-specific GO terms that were significant in at least 15 tissues. Then, since the brain is highly affected by age-related processes and neurological conditions show sex differences, the sex-biased AS events could help explain differential susceptibility to age-related cognitive decline and neurodegeneration. That's why we chose the brain data for detailed analysis.

To improve clarity, we have revised the text to describe the purpose of our analysis in brain rather than other tissues (Page 6-7). We appreciate the reviewer's feedback, and we will consider additional analyses to further explore the sex-biased AS as well as disease risk in other tissues.

(6) The experimental data presented in Figures 5E - I merely demonstrate that estrogen receptor regulates the expression of two splicing factors, SRSF1 and SRSF7, in an estradiol-dependent manner. However, this finding does not support the notion that this regulation actually contributes to sex-dimorphic alternative splicing changes during human aging. Notably, the authors do not provide evidence that SRSF1 and SRSF7 expression changes actually occur in a sex-dependent manner with human aging (in a manner similar to TIA1). As such, this experimental dataset is disconnected from the main focus of the study and does not substantiate the conclusions on sex-dimorphic splicing during human aging. The authors performed RNAseq in wild-type and ER mutant cells, and they should perform a comprehensive analysis of ER-dependent alternative splicing and compare the results with the GTEx data. It should be straightforward.

Thanks for the reviewer's feedback. The main purpose of the analyses in Figures 5E-I was to explore which factors affect the sex-biased expression of splicing factors during aging and substantially regulate alternative splicing (AS). To address the reviewer's concerns, we have included additional analysis and explained the challenge of linking estrogen receptor (ER)-regulated splicing factors to sex-dimorphic AS changes during human aging in specific human cell types.

- As suggested by the reviewer, we first examined the expression changes of SRSF1 and SRSF7 during aging in males and females, like TIA1 in decision-related brain regions (Fig. 5I).
- Secondly, the regulation is based on a highly complex regulatory network involving multiple splicing factors and cell heterogeneity. Due to these complexities, we did not overlap ER-dependent AS changes with sBASEs from GTEx datasets directly. As far as the reviewer is concerned, we supplemented the AS analysis in the GSE89888 dataset (Fig. 5H) and identified

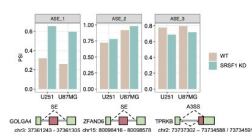
the estrogenregulated AS events mediated by ESR1. We found that ~6% (26/396) of female-specific ageassociated AS events were regulated by ESR1, of which 6 sBASEs can be regulated by femalebiased splicing factors. The low overlaps could be represented by the limited coverage of different RNA-seq datasets and cell types used across these analyses. Notably, the results indicated that only a fraction of AS could be directly accounted for by estrogen via ESR1, suggesting the complexity of transcriptional and splicing regulatory networks during aging.

• Meanwhile, we downloaded independent experimental datasets to discover the regulation by our candidate splicing factors. Due to SRSF1 is identified as a potential regulator of sex-biased splicing, we analyzed RNA-seq data with SRSF1 knock-down (KD) glioblastoma cell lines (U87MG and U251), a type of brain cancer formed from astrocytes that support nerve cells 4. As a result, we indeed found that some sBASEs are regulated by SRSF1 during aging through this experiment using brain cell lines (Author response image 4). Together, these results suggested that some of the SF-RNA regulatory relationships can be observed in another cellular system, further supporting our findings.

Due to the limitations of cell-based models and the complexity in the splicing regulatory network, it is challenging to directly validate aging regulation, particularly between different sexes, based on ER treatments in vivo. However, our findings still provide valuable mechanistic insights into ER-regulated splicing factors, implying their potential role in sex-biased aging.

Author response image 4.

SRSF1 regulations on specific sBASEs using SRSF1 knock-down RNA-seq data in GBM cells. Three examples are shown to be regulated during aging with significant changes between SRSF1 KD vs control in U251 and U87MG cell lines. The splicing diagrams are shown below.



Recommendations for the authors:

Reviewer #1 (Recommendations for the authors):

The authors found that alternative splicing was affected by both sex and age across many tissues, with gene expression differences affected by both parameters only present in some tissues. This trend was consistent when the effects of sex chromosomes were subtracted from the analysis. The effect of aging on differential gene expression and alternative splicing was more prevalent in male than female samples. For analysis purposes, young subjects were deemed to be anyone under 40, and old subjects were over 60 years old. The authors then investigated if specific genes or alternative splicing events were responsible for these effects. Some candidate genes or splicing events were identified but there was little overlap between tissues, suggesting no universal gene or event as a driver of aging. Surrogate variables like the ethnic backgrounds of donors were also investigated. Ultimately the authors found that alternative splicing events showed a stronger sexual dimorphic effect with age than did differential gene expression and that at least for the brain, alternative splicing changes showed a bias for Alzheimer's disease in male samples. This was highlighted by examples of exon skipping in SCL43A2 and FAM107A in males that were associated respectively with plaques and tangles.

The authors go on to identify sexual dimorphic differences in splicing factors in particular brain regions during age. Finally, the authors performed analysis for aging-modulated genes, identifying nearly 1000 across the tissues, nearly 70% of which are sex-specific. Their work suggests that further analysis of these aging-modulated genes could be differentially modulating the transcriptome based on sex. The work is novel and interesting, especially investigating sexual dimorphism in alternative splicing. However, the work is still preliminary, and these assumptions need to be applied to other data sets beyond GTEx for validation as well as some other phenomena that need to be considered. I recommend major revisions to address the points below.

(1) At the beginning of the results section, the authors state that the brain is stratified into four functional regions. It would be useful to explicitly state those four regions in the text at that point.

We agree that specifying these regions early in the text will improve clarity and provide the reader with a clear understanding of the analysis. As the reviewer's suggestion, we revised the Results section (Page 3) to explicitly state the four functional brain regions as follows: 'Due to data sparseness, the brain tissues were recombined into four functional regions (table S1), including hormone- or emotion-related region, movement-related region, memory-related region, and decision-related region (See Methods)'. This ensures that the regions are clearly defined before the subsequent analysis is presented.

(2) The manuscript becomes a bit confusing when the authors shift from all the tissues as a whole specifically to the brain and then back to the larger tissue set to make assumptions. This can be a bit confusing and should be better delineated.

We thank the reviewer and editor for the feedback regarding the transitions between the analysis of all tissues and the brain-specific analysis. In our study, we first conducted a broad analysis of alternative splicing (AS) and gene expression (GE) across all tissues. For the AS analyses, we did sBASEs analysis in all tissues and then focused on specific tissue (i.e., brain) whose splicing changes are functionally enriched with age-related diseases. For the GE analyses, we also analyzed the aging rate across tissues and identified the tissue-specific/shared patterns.

We agree that the shifts of the tissues for AS and GE may cause some confusion, and have made the following revisions to delineate why we focused on different tissues for distinct analyses:

- We have added clear statements to better delineate when we shift focus from the analysis of all tissues to the region-specific analysis and vice versa. For instance, in the Results section (Page 67), we include a transitional phrase: 'Having established patterns across all tissues, we now turn to a more focused analysis to investigate tissue-specific alternative splicing changes.'
- To improve the overall structure, we have reorganized the Results section, adding distinct subheadings for the analysis of all tissues and the brain (Page 6), which should make the transition between these sections smoother and more intuitive for the reader.

We believe that these revisions will make the manuscript's structure clearer and allow the reader to better follow the flow of the analysis and the subsequent conclusions.

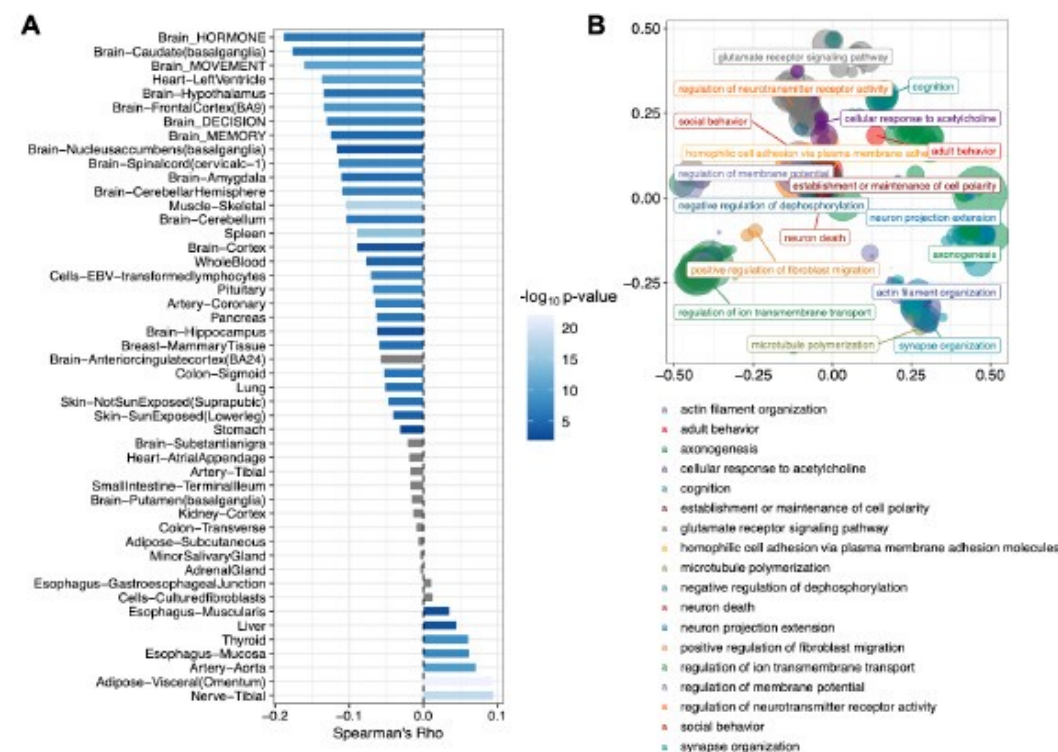
(3) Gene-length-dependent transcription decline (GLTD) is another phenomenon that occurs with aging and is known to be associated with Alzheimer's disease [PMID38519330]. The authors should make some statement if this is present in their dataset and if any sexual dimorphism in tissues is present.

We thank the editors and reviewers for bringing up the possible connection of gene-length-dependent transcription decline (GLTD), which was reported to be associated with both aging and Alzheimer's disease (AD). We appreciate the reviewer's suggestion and have addressed whether GLTD is present in our dataset and whether any sex differences are observed in this context.

We evaluated GLTD using the correlation between gene length with age-associated changes (i.e., the coefficients of the 'age' term in the linear regression model) in GTEx data. We did observe strong evidence of GLTD, particularly in the brain, heart, muscle, pancreas, spleen, skin, muscle, etc (Author response image 5A). In brain, we performed the functional enrichment analysis on the genes with Foldchange > 2 and length > 10⁵ bp (Author response image 5B). We found that these extremely long genes are significantly relevant to synapse and neuron functions. These findings align with previous studies showing that GLTD can occur with aging in the tissues that are relevant to Alzheimer's disease, cardiovascular diseases, and common failures of metabolism (e.g., diabetes) [5,6]. Additionally, it was not a ubiquitous phenomenon across all tissues. The correlations could be positive in tissues like adipose and artery. These findings suggested the GLTD could be varied and tissuespecific in its manifestation during aging.

Author response image 5.

(A) The correlation between gene length and age-associated changes across GTEx tissues in human samples. The correlation tests are evaluated using Spearman's approach. The color bar indicates the -log₁₀ transformed p-values in the correlation test. (B) The results of GO enrichment analysis using the genes with Foldchange > 2 and length > 10⁵ bp. The parent terms calculated by 'rrvgo' with a similarity threshold of 0.9 are shown.

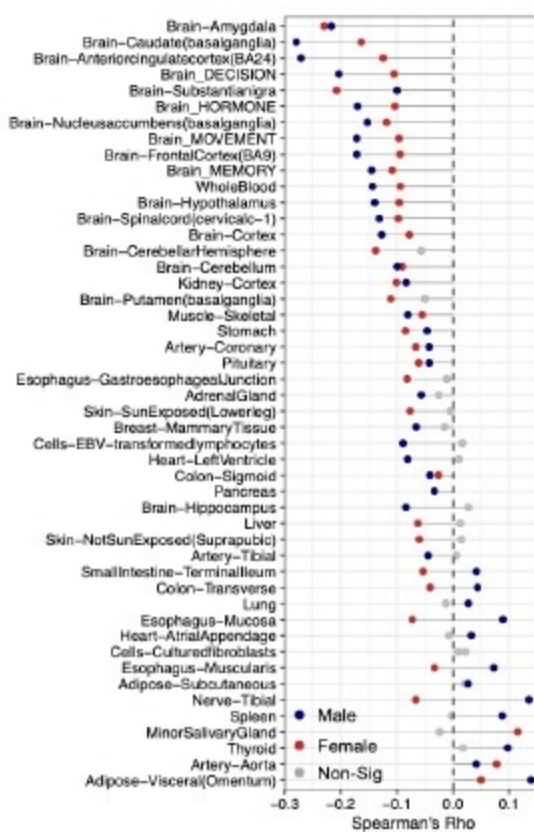


Regarding sexual dimorphism, we conducted this analysis in females and males, respectively (Author response image 6). We found GLTD exists in both females and males in most tissues, such as brain, whole blood, muscle, etc, consistent with the previous results without

considering the sex groups. Interestingly, we observed sexbiased patterns in certain tissues. In particular, the left ventricle, pancreas, and hippocampus showed notable male-biased patterns in the degree of transcriptional decline with gene length, whereas skin, liver, small intestine, and esophagus showed that in females. These findings suggest that GLTD could be relevant to aging and age-related diseases; the levels of expression and sexual dimorphism may vary depending on the tissue type. We hope this clarification addresses the reviewer's concern and provides a more comprehensive understanding of the GLTD and sex differences observed in our dataset.

Author response image 6.

The correlation between gene length and age-associated changes across tissues in females and males, respectively. The correlation tests are evaluated using the Spearman's approach. The red dots indicate the significant correlations in females, while the navy dots show those in males.



(4) Because the majority of this work has been performed in the GTEx dataset, applying this analysis to another publicly available dataset would be useful validation. For instance, the authors have interesting findings in the brain and correlations to Alzheimer's disease. Analysis of an existing RNAseq dataset from Alzheimer's disease patients and controls (with functional outcomes) would provide more evidence beyond the preliminary findings from GTEx.

We appreciate the reviewer's suggestion on the validation of our findings by applying our analysis to independent RNA-seq datasets from Alzheimer's disease patients.

- We have used two Alzheimer's disease datasets, GEO and ROSMAP, to investigate the correlation between aging and Alzheimer's disease (AD) and included these analyses in our

study (Fig. 4B-C and Figure S8C).

- In the Results section (Page 7), we have presented the results of this validation, where we identified correlations between sex-biased aging-related splicing changes and AD-related changes. These findings support the conclusions from the GTEx dataset and further strengthen the relevance of our results to AD.

As suggested, we have updated the manuscript to more explicitly highlight this validation in the Discussion section (Page 12), noting: ‘We further validated our findings using Alzheimer’s disease dataset, ROSMAP, where we observed consistent correlations between aging-related splicing changes and Alzheimer’s disease-related changes, providing additional evidence for the robustness of our results.’

Reviewer #2 (Recommendations for the authors):

(1) In the text (Introduction and Discussion), the authors mention analyzing 54 tissues, the abstract states 35 tissues, Table S1 lists 48, and Figure 2A-B shows 33. Could the authors please clarify exactly how many tissues they used? I am also confused by the sample numbers in Table S1. For example: for adiposesubcutaneous tissue, the total number of females is listed as 218 but the sum of young and old females is only 110. Does this mean some samples were excluded? What is the exclusion criterion?

We thank the reviewers and editors for pointing out the discrepancies regarding the number of tissues analyzed and the sample numbers in Table S1. We appreciate the opportunity to clarify these points:

Number of tissues analyzed:

- We downloaded and analyzed 17,382 samples in 54 tissues from GTEx in total (31 tissues and 13 brain regions), as mentioned in the Results, Methods, and Discussion sections. Table S1 lists 48 tissues (31 tissues, 13 brain regions, and 4 merged brain regions), which include a refined classification of the tissues we analyzed, accounting for the variations in brain region categorization in the dataset.
- The discrepancy also arises from the different sample size cutoffs in specific analyses. For pcSVR analysis (Figure 2A-B), we did the subsampling for the permutation analysis for certain key findings, so we filtered a subset of 33 tissues (29 tissues and 4 merged brain regions), which included at least 3 samples in each age group in females or males.
- To resolve this, we have clarified the total number of tissues analyzed and aligned the numbers across the manuscript. In the revised manuscript, we now explicitly state in both the Abstract and Methods sections that 54 tissues were analyzed in the context of this study. We added a note in Methods to clarify that 35 tissues are 31 tissues and 4 merged brain regions (Page 16). In Figure 2A-B, we clarified that the 33 tissues are filtered due to the usage in this analysis (Page 17).

Sample numbers in Table S1:

- Regarding the sample sizes of age groups, the discrepancy occurred due to the classification of the age groups. We classify the samples into three: Young, Middle, and Old, as mentioned in the Results section (Page 4).
- Additionally, we excluded the sample sizes in 13 single brain regions. We aligned the total tissue number to 35 with our texts.

We hope this resolves the confusion regarding the number of tissues and the sample sizes used in the analysis. These clarifications have been incorporated into the revised manuscript

to ensure consistency.

(2) Was post-mortem interval (PMI) or manner of death considered in the model? For example, traumatic death may have major consequences on gene expression. Similarly, a few tissues have low sample numbers, for example, kidney cortex and brain. The pooling of brain samples is explained and the kidney cortex is excluded, so why is it listed in Table S1?

Thank you for raising this important point regarding the potential impact of post-mortem interval (PMI) and manner of death (DTHMNNR) on gene expression. We carefully considered both factors as potential confounders in our analysis.

Specifically, to evaluate their impacts, we calculated the correlations between the coefficients of PMI or manner of death, with the confounding factors. Our results showed that PMI and DTHMNNR are significantly correlated with the covariates in most tissues, suggesting that their effects could be effectively regressed in our model (Figure S4). As we have mentioned in Figure S4 and Author response image 1, we conducted a differential analysis that incorporated PMI as a covariate in the regression models and re-evaluated the age- and sex-related transcriptomic changes to address this concern. The high correlations showed the minor effect size of PMI when including the covariates in the model. As suggested by the reviewers and editors, we have now included this correlation analysis in Figure S4C-E and updated the text in the results section (Page 5).

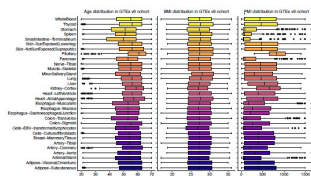
Additionally, as the responses above, Table S1 provides the general sample sizes of all GTEx tissues without filtering. We have modified the table to include a total of 35 tissues, including 31 non-brain tissues and 4 brain regions.

(3) It might be important to show a simple visual of cohort details such as age ranges, sexes, ethnicities, PMIs, etc.

To address this, we added summary figures to illustrate the distributions of key demographic variables, including age, sex, BMI, ethnicity, post-mortem intervals (PMIs), and manner of death (DTHMNNR) (Author response image 7 and Author response image 8). This will provide readers with a clearer overview of the dataset composition and potential covariates affecting the analysis.

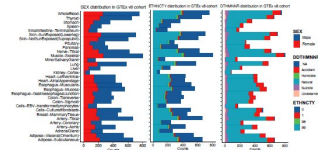
Author response image 7.

Age (left panel), BMI (Body Mass Index) (middle panel), and PMI (Post-Mortem Interval) (right panel) distribution in GTEx v8 cohort.



Author response image 8.

Sex (left panel), ethnicity (middle panel), and manner of death (DTHMNNR) (right panel) distribution in GTEx v8 cohort.



(4) Since this study is highly correlative, it is impossible to determine if the findings hold true without an independent cohort validation or experimental validation. They used the ROSMAP cohort for AD samples, and some splicing factors regulation but the generalizability to the age and sex effects have not been independently tested.

The reviewer raises an important point regarding the independent validation of sex- and age-associated splicing changes associated with AD. We used GTEx primarily because it includes approximately 17,000 RNA-seq samples across multiple human tissues, making it the most comprehensive public resource for studying population-level differences in age and sex. In particular, its large-scale brain samples provide a unique opportunity to analyze transcriptomic changes in sex-dimorphic aging.

We understand the reviewer's concern that our findings are mainly supported by correlative evidence, which could be affected by dataset-specific biases. However, there are several technical issues in crossvalidation with transcriptomes across different datasets, including limited comparability due to cell type heterogeneity, postmortem artifacts, and sequencing biases.

Specifically, GTEx data is bulk RNA-seq that does not capture cell-type-specific transcriptomic changes. Given the cellular complexity of the brain and other tissues, observed differences in gene expression and splicing may be influenced by shifts in cellular composition rather than intrinsic transcriptional regulation. For example, we compared our results from GTEx whole blood with the analysis using an external dataset from Peripheral Blood Mononuclear Cells (PBMCs) provided by Shen et al. (2024) [3] (Author response image 2). We observed limited overlap in differentially expressed genes between these datasets (probably because the whole blood contains diverse immune cell populations), highlighting the challenges in cross-dataset validation due to differences in tissue composition and sample processing.

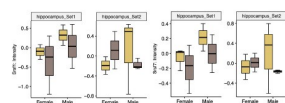
Therefore, we applied surrogate variable analysis (SVA) to minimize technical and biological confounders. This approach helped reduce biases from genetic background to hidden batch effects, including postmortem artifacts, sequencing biases (Figure S4), and other covariates. This approach could help us identify whether sex-biased splicing events are biologically meaningful rather than technical artifacts.

In addition, to address the reviewer's concern on the splicing factor regulation, we managed to find a dataset in decision-related brain regions. Due to the limitation of human brain data covering different age and sex groups, we used mouse hippocampus datasets, including young and old, as well as female and male groups [7]. The analysis of protein levels from MS data identified sex-biased age-associated splicing factors, including Srsf1 and Srsf7. We found that the changes are consistent with the findings from GTEx (Author response image 9), aligning with our sex-biased splicing factor expression during aging in the same region of the human brain. This cross-species consistency supports the robustness of our findings in human brain aging.

Author response image 9.

Protein levels of some male-specific splicing factors in human hippocampus quantified using MS data. The Y-axis shows the protein intensity. Different facets mean different sample batch

sets. The yellow boxes indicate the protein levels in the young group, while the brown boxes indicate those in the old group.



In summary, despite the inherent limitations of RNA-seq studies in sex- and age-related transcriptomics, we have made our best efforts to address these concerns through comparisons with external datasets, statistical corrections, and validation using proteomic data. We appreciate the reviewer's feedback and include additional discussion on these points (Page 13).

(5) Are AS predictions from short-read data accurate enough to make the predictions the authors report?

The reviewer is correct that the short-read sequencing has inherent limitations in reconstructing full-length isoforms. However, the higher sequencing depth for short reads makes it a better choice in quantifying the relative change of each AS event across different conditions. As a result, short-read data are extensively used in the splicing field to quantitatively measure the AS changes. For this reason, we focused on the levels of alternative splicing events, rather than the quantification of full-length isoforms. We used a series of stringent filters in our analyses to increase the reliability of our results.

Specifically, we filtered the read counts of the junction read counts (JC) of most differential AS events that were higher than 10, as mentioned in the Methods section. Also, we used our GPU-based gene expression quantification tool, Paeon, which performed better in cross-validation with quantitative RT-PCR results. The results of Paeon are consistent with other pipelines. We cited an updated version of Paeon that included the comparison with other tools in analyzing AS for consistency. The manuscript on the new Paeon version is being reviewed in another journal, and we included the PDF of that manuscript (Fig. 3 in the Paeon manuscript) in the revised documents.

(6) Along the same lines, the finding that male age-related AS events are linked to Alzheimer's disease somewhat contradicts epidemiological studies that show that even after adjusting for age, women still have a greater risk of developing Alzheimer's than men. The authors show a significant overlap with AD GE events in females but don't explain the discrepancy.

We appreciate the editor's comment regarding these discrepancies with the epidemiological studies. Previous studies suggested that the disease manifestations of Alzheimer's Disease (AD) showed sex differences in AD phenotypes, including cognitive decline and brain atrophy [8]. The analyses on the sex/age effect of AD are indeed pretty complex, depending on the molecular criteria (GE or AS vs epidemiological data) in distinct studies, probably due to the difficulty in capturing how environmental exposures interact with biological pathways. We hope to bring up three related points regarding this concern, which were also discussed in the revised manuscript.

- As we have mentioned in the Discussion section, an early study investigated the relationship between age, sex, and cognitive function in a large cohort of 17,127 UK Biobank participants [9]. Their study highlighted more apparent age-related changes in cognitive function among men, suggesting a potential vulnerability of men to cognitive decline with age. Their main conclusion is consistent with our findings.

- While men and women can both suffer from Alzheimer's disease, women are more likely to be diagnosed, possibly due to longer lifespans and potential differences in brain structure or other factors. Although women exhibit a higher overall risk of AD, they may also have distinct molecular compensatory mechanisms that influence disease progression.
- To avoid the age effect, in our AD datasets, including ROSMAP, we filtered the samples over 90 years old to match the number of both sexes and the age distribution between the AD and control groups. Our analysis avoided the age biases in comparing AD and control, suggesting the crucial roles of sBASEs in AD during male aging.

Moreover, for gene expression (GE), we showed distinct patterns of AD-related genes in females with AS. These two molecular processes do not necessarily have the same functional impact. AS changes may precede or contribute to disease onset in different ways compared to GE alterations. Our study came up with the underlying mechanisms linking cognitive disorders and alternative splicing (AS) at a higher molecular resolution.

(7) Could the authors explain which sBASE subset they used for their random forest prediction model and what was the rationale?

We are sorry for missing the details in selecting sBASEs (sex-biased age-associated splicing events) for the random forest prediction model. We specifically used sBASEs that exhibited specific sex-biased changes in splicing associated with aging. This subset of sBASEs was chosen in terms of those that could also be detected in the ROSMAP AD dataset due to different sequencing depths or technical biases across datasets. These sBASEs were further input to a prediction model with the feature selection algorithm RFE, and then evaluated their contributions. In the revised manuscript, we added the details of this selection in the Methods (Page 7).

(8) The breakpoint analysis is particularly interesting. Can this be speculated to correlate with the recent non-linear multi-omic aging patterns observed by Shen et al in Nature Aging?

Thank you for highlighting the interesting aspects of our breakpoint analysis and suggesting its potential correlation with the non-linear aging patterns observed by Shen et al.

Shen et al. observed two prominent crests around the ages of 45 and 60 using omics data. Similarly, we also identified the non-linear aging patterns with two breakpoints in our analysis. However, there are some notable differences in specific breakpoints between these two studies, resulting from the breakpoint definition, as well as the sample preparations. According to the response in Author response image 2, the differences come from the following aspects:

The definition of breakpoints vs crests:

- Crests represent age-related molecular changes at each time point across the human lifespan. They indicate the number of molecules that are differentially expressed during aging ($q < 0.05$), without considering individual expression levels.
- Our breakpoints, in contrast, are identified after filtering the chronological trends based on the expression levels and calculating the rate of change at each age point using sliding windows. Breakpoints are defined as local maxima where the distance to the nearest minimum, relative to the global maximum, exceeds 10%. We indeed found some local wide peaks around 60 in some tissues, shown in Figure S10, however, we excluded these due to our strict cutoffs.

The sequenced biosamples:

- Whole-blood vs Peripheral Blood Mononuclear Cells (PBMC): As mentioned in previous responses, in GTEx, whole blood samples from donors were sequenced, whereas their study used PBMCs. Whole blood contains all blood components, including red blood cells, platelets, granulocytes (e.g., neutrophils), lymphocytes, and monocytes, while PBMCs only represent a subset of white blood cells, primarily consisting of lymphocytes (T cells, B cells, NK cells) and monocytes, excluding granulocytes and erythrocytes. Gene expression changes observed in whole blood capture the contributions from neutrophils and other granulocytes, which are absent in PBMC analyses (as shown in Figure S11C and Author response image 2). Additionally, whole blood can serve as a readily accessible biomarker source for testing age-related diseases without the need for cell separation, making it a more practical option for clinical applications.

- For both studies, we share a tissue, which is skin, we looked at the non-linear changes during aging and found the same two breakpoints: 43 and 58.

Sex-specific analysis in females and males:

- The main object of our analysis is to compare the differences in aging rates between sexes. Notably, the identified breakpoints may differ when sex effects are not taken into account, highlighting the importance of analyzing males and females separately.

We have added the following statements to further clarify this connection: ‘Our analysis observed the nonlinear aging patterns with two breakpoints, which is consistent with recent findings (Nature Aging, 2024), with differences in specific age points due to the sex differences as well as tissue diversities.’ (Page 14), and ‘These breakpoints could represent key junctures in the aging process that align with the non-linear patterns of aging and disease progression.’ (Page 15)

(9) Minor - the authors should refer to figures in the Discussion. They do so in some cases but this needs to be more extensive.

Thank you for pointing this out. In response, we have reviewed the Discussion section and added references to relevant figures where appropriate. In the section discussing the discrepancies between the profiles of GE vs. AS, we now refer to Figure 3 to highlight the earlier onset of different transcriptomic resolutions (Page 12); When describing the sex-specific age-associated AS changes and their associations with Alzheimer’s disease, we have added references to Figure 4 (Page 12); In the discussion of estrogen-mediated regulation of splicing factors, we have referred to Figure 5A, which detail the construction of RBP-RNA regulatory network integrating multi-dimensional data obtained through several orthogonal state-of-the-art approaches (Page 14).

Reference:

- (1) Ferreira, P.G. et al. The effects of death and post-mortem cold ischemia on human tissue transcriptomes. *Nature communications* 9, 490 (2018).
- (2) Wucher, V., Sodaei, R., Amador, R., Irimia, M. & Guigó, R. Day-night and seasonal variation of human gene expression across tissues. *PLoS Biology* 21, e3001986 (2023).
- (3) Shen, X. et al. Nonlinear dynamics of multi-omics profiles during human aging. *Nature aging*, 116 (2024).
- (4) Zhou, X. et al. Splicing factor SRSF1 promotes gliomagenesis via oncogenic splice-switching of MYO1B. *The Journal of clinical investigation* 129, 676-693 (2019).

- (5) Soheili-Nezhad, S., Ibáñez-Solé, O., Izeta, A., Hoeijmakers, J.H. & Stoeger, T. Time is ticking faster for long genes in aging. *Trends in Genetics* 40, 299-312 (2024).
- (6) Brouillette, M. Gene length could be a critical factor in the aging of the genome. *Proceedings of the National Academy of Sciences* 121, e2416630121 (2024).
- (7) Keele, G.R. et al. Global and tissue-specific aging effects on murine proteomes. *Cell reports* 42(2023).
- (8) Ferretti, M.T. et al. Sex differences in Alzheimer disease—the gateway to precision medicine. *Nature Reviews Neurology* 14, 457-469 (2018).
- (9) Foo, H. et al. Age-and sex-related topological organization of human brain functional networks and their relationship to cognition. *Frontiers in aging neuroscience* 13, 758817 (2021).

<https://doi.org/10.7554/eLife.102449.2.sa0>