Pan-tissue Transcriptome Analysis Reveals Sex-dimorphic Human Aging

Authors
Siqi Wang [1], Danyue Dong [1], Xin Li [1], Zefeng Wang [1,2,3]*

Affiliations
[1] CAS Key Laboratory of Computational Biology, Shanghai Institute of Nutrition and Health, University of Chinese Academy of Sciences, Chinese Academy of Sciences, Shanghai 200031, China
[2] Bio-Med Big Data Center, Shanghai Institute of Nutrition and Health, University of Chinese Academy of Sciences, Chinese Academy of Sciences, Shanghai, 200031, China
[3] School of Life Science, Southern University of Science and Technology, Shenzhen, Guangdong, 518055, China

* Corresponding to: wangzefeng@picb.ac.cn

Abstract

Complex diseases often exhibit sex-dimorphism in morbidity and prognosis, many of which are age-related. However, the underlying mechanisms of the 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 females-biased events are often regulated by several sex-biased splicing factors that may be controlled by estrogen receptors. Breakpoint analysis showed sex-dimorphic aging rates that 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 recent 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 were 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 mechanism 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 development 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 cancer [29]. Since most introns are co-transcriptionally spliced, AS is also affected by various factors that regulate transcription, such as promotor activity, Pol II elongation, and chromatin modification/remodelling [28]. Recent studies reported that AS regulation is closely associated with sex [30] and age [31-33]; 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 project [21] 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). 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 study [34], 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 [35,36], 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 ratio [34], 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.

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

151 coefficient term (see Materials and Methods, Equations 4 and 5). In each tissue, we
152 identified the sex- or age-differential genes and AS events (fig. S3, A and B).
153 Consistent with previous reports [34,37,38], the breast tissue contains the largest amount
154 of sex-differentially expressed genes, while the age-biased genes were widely
155 discovered in brain regions and artery-related tissues. Interestingly, in most tissues,
156 there were no significant overlaps between age-differential genes and AS events (fig.
157 S3A, hypergeometric test p-value < 0.01). In contrast, we found significant overlaps
158 between sex-differential genes and AS events in most tissues (fig. S3B).
159 As a control, we constructed a separate regression model to evaluate the
160 contribution of sampling biases, including sample sizes and the total number of genes
161 or AS events detected in each tissue (Equation 6). We found that the sample size did
162 not significantly affect the number of identified sex-differential or age-differential
163 genes/AS events (fig. S3, C and D), suggesting our analyses are not affected by
164 sampling biases. Furthermore, using the coefficient of the age-by-sex term in this
165 model allowed us to identify thousands of genes and AS events affected by the
166 functional interactions between sex and age (i.e., the sex effects on GE/AS depend on
167 the age, *or vice versa*) (fig. S3E), suggesting extensive sex-dimorphic changes of
168 transcriptome during aging.
169 A recent study reported the environmental effect on human genetic background
170 during aging [39], which should be considered when analyzing large datasets from
171 different populations. In the linear regression model, we included the surrogate
172 variables (SV) as confounders to control for the effect of genetic backgrounds
173 (Equations 4 and 5). Using the differential GE analysis in whole blood as an example
174 (other tissues have the same trend), we evaluated the correlation between SVs with
175 donors' ethnicity using Point Biserial Correlation, and found that many SVs showed
176 significant correlations with the donors' ethnical background (fig. S4A), suggesting
177 that these SVs could effectively reflect the variables of genetic background.
178 Additionally, we calculated the correlation between SVs with the top 5 principal
179 components as judged by whole genome sequencing (WGS) data (fig. S4B), and
180 found roughly the same set of SVs also showed a significant correlation with PC1.
181 Therefore, the SVs in the linear regression models could effectively control the
182 cofounders of human genetic regulation.
183 To further evaluate the robustness of the differential analyses, we performed
184 detailed assessments of the linear regression models and the selection of GE/AS
185 cutoffs using the decision-related brain region as an example. We calculated the
186 fraction of original differential genes/AS events identified in each permutation
187 iteration, and the false discovery rate (FDR) for each gene/AS event using the
188 permutation analysis in the linear regression model (see Materials and Methods).
189 Briefly, we generated the permuted data 1,000 times by randomly shuffling the group
190 labels (sex and age) with the same sample sizes as the original labels. The distribution
191 of the p-values of sex/age-differential genes/AS events identified based on the
192 original and each shuffled phenotype data were evaluated (fig. S4, C and F). We first
193 observed low fractions of sex/age-differential genes/AS events in most iterations
194 under our cutoffs (fig. S4, D and G). We found that FDRs of most sex/age-differential

genes/AS events were not more than 5% using the linear regression model (fig. S4, E and H), which is much lower than the FDR of DESeq2 and edgeR [40]. 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. S5, 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 expression and age-differential AS under most combinations of TPM and JC cutoffs in decision-related brain region (fig. S5C), 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

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 as judged by hypergeometric tests (p-values < 0.01, 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. S5D). 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

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. S6, table S4). Interestingly, 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. S7A). Because previous studies suggested that the disease manifestations of Alzheimer's Disease (AD) showed sex differences in cognitive decline and brain atrophy [4], 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., $PSI_{AD} - PSI_{Ctrl}$) 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=6.6×10^{-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. S7B), supporting a recent finding that females are more vulnerable to AD during aging due to the post-menopause activation of C/EBPβ–AEP/δ-secretase pathway by sex hormone changes [41].

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, $p < 2.2\times10^{-16}$) than such correlation across the sBASEs in females (R=0.371, $p < 2.2\times10^{-16}$) (fig. S7C), 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 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. S7D, 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 [27,42]. Dysfunction of the SFs has been reported as a hallmark of aging, presumably by regulating the genes in cellular senescence during aging process [32]. 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 [43]. 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 [36,44]; 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

326 female-specific age-associated SFs were significantly higher than the median value of
327 the age-associated SFs common in both sexes (Fig. 5E; p = 0.032), suggesting that
328 ESR1 was more likely to bind the promoters of the female-specific age-associated
329 SFs. In addition, the age-associated decline of ESR1 was more substantial in females
330 than males (Fig. 5F), and some of the female-specific age-associated SFs (e.g., SRSF1
331 and SRSF7) could be stimulated by propyl pyrazole triol, an agonist of ERα (Fig. 5G)
332 [45].
333 These observations were further confirmed by an independent experiment where
334 the wildtype and two constitutively active mutants (Y537S, and D538G) of ESR1
335 were engineered into MCF-7 cells using CRISPR-Cas technology [46]. The resulting
336 cells were treated with estradiol or vehicle controls and subjected to transcriptome
337 profiling by RNA-seq, and the data were reanalyzed using the same pipeline. We
338 found that the basal levels of SRSF1 and SRSF7 were significantly higher in the cells
339 with constitutively active ESR1 (Fig. 5H, comparing samples treated with vehicle
340 controls), and SRSF1 and SRSF7 were further induced by the estradiol treatment only
341 in cells with wildtype ESR1 but not in cells with mutated ESR1 (Fig. 5H, compare the
342 samples treated with estradiol *vs.* control in different cells), indicating that the
343 activation of ESR1 pathway can promote expression of SRSF1 and SRSF7. The
344 similar changes were also found in other splicing factors including HNRNPA2B1 and
345 HNRNPC (fig. S8A). As expected, the expression levels of SRSF1 and SRSF7
346 showed an obvious decrease during female aging (Fig. 5I).
347 To further confirm our findings in different tissues, we analyzed the RNA-seq
348 data with ER knockout and estrogen treatments in mouse brain arcuate nucleus
349 (ARCs) [47] and found a consistent trend in the expression changes for splicing factors
350 Srsf1, Srsf7, and Hnrnpa2b1 under estradiol (E2) regulation (fig. S8B). Specifically,
351 these splicing factors were activated by estrogen in wild-type mice but exhibited
352 reduced expression in ER knock-out mice. The direct binding of estrogen was also
353 examined through integrating the ESR1 ChIP-seq data [48]. As expected, a major ESR1
354 binding peak was found in the promoter regions of Srsf7, which was increased by
355 estrogen treatment (fig. S8C).
356 Collectively these analyses across multiple datatypes suggested that the female-
357 specific SFs (such as SRSF1 and SRSF7) are more sensitive to the estrogen changes
358 *via* ESR1-mediated pathway, leading to extended sex-biased AS regulations during
359 aging.
360
361 **The aging rates of GE show sex-dimorphism during aging across multiple tissues**
362 In the earlier results (Fig. 3A), we found a large proportion of age-associated
363 genes are common in both sexes. Therefore, we asked whether the levels of GE
364 change at the same pace during aging in females and males. The age-associated genes
365 ubiquitously changed in both sexes were analyzed for consistency. We first used the
366 Autoregressive Integrated Moving Average model [49] to capture the age-associated
367 genes with significant chronological changes beyond random fluctuations, which are
368 the main contributors to transcriptome dynamics during aging. We further quantified
369 the GE changes between adjacent age windows using breakpoint analysis [16] (Fig. 6A,

370 see Materials and Methods). Such analyses were conducted across all tissues in both
371 sexes (fig. S9A), with two tissues shown in Fig. 6B. A higher value on the Y-axis
372 (mean -log P) indicated a larger GE change at the given age, and thus the peaks were
373 regarded as aging "breakpoints" that represent the key time point of transcriptome
374 aging. Compared to females, males showed significantly larger GE changes during
375 aging in most tissues (fig. S9A), except transverse colon, fibroblast cells, and whole
376 blood (fig. S9B). The number of breakpoints also showed sex-dimorphism in several
377 tissues. For example, in sun-exposed skin and colon, males showed two main
378 breakpoints around ages 35 and 50, whereas females only have one breakpoint at age
379 ~45 (fig. S9A).
380 Moreover, we found that the major aging breakpoints (i.e., breakpoints with the
381 largest GE changes) occurred earlier in males than females across most tissues, and
382 the magnitudes of changes were also more prominent in males (Fig. 6C), suggesting
383 that males may age faster with a larger age-associated GE change. In addition, we
384 conducted breakpoint analysis with all chronological genes in both sexes and found
385 similar results when comparing the breakpoints in males and females (fig. S9C). We
386 also performed a similar analysis using chronological sex-biased age-associated AS
387 events (sBASEs) and all chronological AS events in females and males separately,
388 and again found an earlier and more obvious breakpoint (fig. S9, D and E). In
389 summary, these results uncovered an earlier aging and a higher aging rate in males at
390 molecular levels, again implying that males could be surprisingly more vulnerable to
391 aging.
392
393 **Aging contributors in separate sexes show distinct functions**
394 To further examine how the individual gene affected aging rates, we designed a
395 disturbance analysis to identify aging-modulated genes (AMGs) (Fig. 6D).
396 Specifically, we randomly removed 20% of chronological genes and repeated the
397 breakpoint analysis with the remaining genes to test if the breakpoints were
398 significantly altered (see Materials and Methods). This process was repeated 200
399 times, and the genes enriched in the 200 random samplings (judged by Fisher exact
400 test $p < 0.05$) were defined as AMGs that serve as the key contributors to molecular
401 aging.
402 We identified a total of 1035 AMGs across different tissues (table S6), around
403 70% of which (723 in 1035) were sex-specific (Fig. 6E). Functional enrichment
404 analysis suggested that the female-specific AMGs were significantly enriched in
405 cellular respiration and NADH dehydrogenase pathways (Fig. 6F), which is consistent
406 with an earlier report of their roles in human aging and longevity [50]. Such functional
407 enrichment was not found in the male-specific AMGs (Fig. 6G), implying that female
408 aging may be more sensitive to mitochondrial and oxidative functions. Intriguingly,
409 we also found many male-specific AMGs relevant to myeloid leukocyte
410 differentiation, phagocytosis, and macrophage apoptotic processes, implying a sex-
411 differential role of the immune system during aging.
412 To determine the potential mechanism of the sex-dimorphic aging rate, we
413 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. S10, A and B). These genes were functionally enriched in neutrophil and granulocyte migration and RNA splicing (fig. S10C). Moreover, we also examined sex-specific AMGs and found very distinct expression patterns (fig. S10, 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. S10D, 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 [51,52]. 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, highlighting the contribution of splicing in sex dimorphism during human aging at an isoform-specific resolution.

Most studies showed that females have a higher risk of AD due to the post-menopause changes of sex hormones [41], however, a higher incidence rate of mild cognitive impairment in males is also reported [53]. 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 [54]. Therefore, it's intriguing to further study the associations between the sex-biased functional property and molecular changes. At the splicing level, we surprisingly observed a male-biased association between age-associated AS changes and AD. 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. 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

476 expectancy in males [12]. Furthermore, we identified many sex-specific AMGs with
477 different biological functions, such as the macrophage and mitochondria functions
478 (Fig. 6, F and G). These genes tend to be regulated by the dominant sex hormones in
479 each sex. The investigation of the aging rate provided insights into the intervention of
480 the aging process or age-related diseases, and to some extent made available for
481 identifying therapeutic targets at specific age points. It is worth noting that our
482 analysis of the aging rate in this section does not include the data from brain tissues
483 due to the sample sparsity in selecting parameters.

484 Our analyses also have some technical limitations, the biggest of which may be
485 caused by the cell type and sample heterogeneity between different donors of GTEx
486 datasets. Since GTEx dataset is the most comprehensive dataset with both
487 transcriptome and whole genome sequencing data, it's a good public resource for sex,
488 age, and age-by-sex interaction analysis within a healthy context. Compared with the
489 earlier studies that used either microarray or other RNA-seq data to analyze the effect
490 of age or sex on brain transcriptome, our main findings supported the main conclusion
491 of these reports, including the diverse gene profiles across brain regions during aging,
492 as well as more vulnerability in males on a global scale [17,18]. In addition, GTEx has a
493 very strict selection of healthy donors to reduce biases from different diseases [55],
494 which is a unique feature compared to other larger cohorts that often contain patients
495 of various diseases that may affect the analyses at the molecular level.

496 The second limitation is the existence of potential confounding factors (like batch
497 effect, genetic background, etc.) that are often associated with big data analyses.
498 However, when such batch effects are entangled with the effect from distinct sex/age
499 groups, it becomes technically challenging to effectively remove such factors without
500 canceling out the contribution of sex/age, which is a common technical difficulty
501 reported also in other analyses [56]. We take several measurements to reduce the effect
502 of confounding factors. First, we utilized the original TPM/PSI data to aggregate
503 sex/age effects and conducted a permutation analysis to assign an empirical *p*-value to
504 each pcSVR. This approach provided a robust statistical framework for evaluating the
505 significance of sex/age effects while accounting for potential batch effects. Secondly,
506 in the context of the differential analysis, we preserved all possible confounders and
507 identified changes that were more obvious than the effects of these confounders, a
508 strategy that reduces false positive discoveries originating from potential batch
509 effects. Another technical limitation is related to the small number of samples in
510 some tissues, and thus we either omitted them from aging analyses (like pituitary,
511 bladder, and kidney) or combined similar tissues (like brain regions) for detailed
512 analyses.

513 Moreover, we derived the AS regulatory networks using the multi-dimensional
514 large-scale data from highly proliferative and easily cultivable cancer cell lines, which
515 may not be an ideal model for the post-mitotic cells in human tissues. However, we
516 dissected the RBP-RNA regulatory network by integrating muti-dimensional data
517 obtained through several orthogonal state-of-the-art approaches, including the eCLIP-
518 seq, *in vitro* RNA affinity evaluation, RNA-seq with RBP depletion, and chromatin

519 association [57]. In addition, the RBP-RNA binding relationships seem to be universal,
520 and the distinct splicing outcomes are mainly determined by the tissue-specific
521 expression and activity of different splicing regulatory factors. Therefore, it is
522 possible to use the data in cancer cells to derive a connectivity map of RBP-RNA, and
523 then use the tissue-specific expression of RBPs to refine the regulatory map.

524 It was reported that there is only a moderate correlation between mRNA and
525 protein abundances for many human genes, mainly because of the variation in
526 translation efficiency of different mRNAs and the additional regulation in protein
527 degradation. Here we constructed the SF-RNA regulation network with the shRNA-
528 seq data that represents the reduction of specific splicing factors in both RNA and
529 protein levels. Such data is more reliable than the correlation of cell-specific
530 expression of RNA and protein, and thus it is feasible to evaluate the consequences of
531 gene silencing by looking at the RNA level. To minimize tissue biases in the RBP-
532 RNA regulation network, we incorporated 3-step processes (see Materials and
533 Methods) including the restrictions of the correlations between RBP expression and
534 AS events within each tissue. Additional large-scale experiments are needed in the
535 future to globally profile the physical and functional interactions between RNA and
536 RBP with a higher accuracy from different tissues.
537 Mechanistically, we found a prominent role of sex hormones in regulating sex
538 dimorphic aging or age-related diseases. Using a range of RNA-seq and ChIP-seq
539 data, we uncovered a type of female-specific age-associated AS regulation (Fig. 5A),
540 which may be achieved partially through the estrogen-mediated regulation of female-
541 specific splicing factors (Fig. 5, E and I). We further quantified chronological
542 changes of transcriptome dynamics using breakpoint analysis, and observed the sex-
543 dimorphic aging rates across multiple tissues. The aging breakpoints are largely
544 consistent with the changes in circulating sex hormones during aging. The level of
545 testosterone starts to decrease at the age 35–40 in males, while in females, estrogen
546 level starts to decrease at a later age (i.e., ~50 years old) [36,44]. Furthermore, we
547 identified a series of female-specific and male-specific aging-modulated genes
548 (AMGs), whose functions are consistent with the reported roles of sex hormones. For
549 example, estrogen was reported to participate in the antioxidant system *via* NADPH
550 [36,44], which is functionally enriched in female-specific AMGs (Fig. 6F). Conversely,
551 androgen was reported to affect the inflammation in immune response [36,44], which is
552 enriched in the male-specific AMGs (Fig. 6G). Consistently, we found that the sex-
553 specific AMGs were more likely to be regulated by androgen receptor in males or
554 estrogen receptor in females (Fig.6H). However, the regulation of sex-biased aging is
555 probably more complicated. Other factors, including mitochondrial functions, genes
556 encoded by sex chromosomes, non-coding RNAs, and transposable elements, may
557 also contribute to the sex-dimorphism during aging [12,14,58,59]. In addition, it is possible
558 that the genetic variants, including expression QTL (eQTLs) and splicing QTLs
559 (sQTL), can affect GE or AS in a sex-biased or age-associated fashion [22,39,60]. Future
560 studies of these factors will improve our understanding of the intricate regulation of
561 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 [61-63], transcriptome [64], circulating immune proteins [65], human gut microbiome [66], and multi-omics features [67]. 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 [68]. Therefore, we speculate that constructing separate aging clocks in distinct sexes will help model the progression of age-related diseases accurately.

572 **Materials and Methods**

573 **Transcriptome quantifications and data processing**

574 We used the transcriptome data from Genotype-Tissues Expression (GTEx)
575 project version 8.0, which contains 17,382 samples in 54 tissues (including 13 brain
576 regions and two cell lines). We downloaded the RNA-seq bam files and the phenotype
577 information from dbGaP (study accession: phs000424.v8.p2; table accession:
578 pht002743.v8.p2.c1). The tissues with a small number of samples and specific to only
579 a single sex (such as ovary and testis) were removed for a better comparison. In
580 addition, for each donor, we combined 13 brain regions into 4 main functional
581 regions, including hormone- or emotion-related region (amygdala, anterior cingulate
582 cortex, and hypothalamus), movement-related region (cerebellar Hemisphere,
583 cerebellum, and spinal cord), memory-related region (caudate, hippocampus, nucleus
584 accumbens, putamen, and substantia nigra) and decision-related region (cortex, and
585 frontal cortex). As a result, for the N donors that each have 13 brain regions (N×13
586 sample matrix), we merged the data into 4 functional regions to generate an N×4
587 sample matrix in the downstream analyses. In total, 16,202 samples in 35 tissues or
588 tissue groups from 948 donors passed these filters for further analysis. The numbers
589 of samples in each tissue are shown in table S1.

590 GE and AS quantifications were performed by Paean [69], a parallel computing
591 system on the GPU-CPU platform with high computational efficiency for super-large
592 datasets. We calculated TPM (transcripts per million) for protein-coding genes and
593 PSI (percent of spliced-in) values for AS events, including skipped exon (SE),
594 alternative 5′ splice site (A5SS), alternative 3′ splice site (A3SS), and retained intron
595 (RI). Genes with low expression (average TPM < 1) were removed. We also filtered
596 the AS events by first removing the samples with missing PSI values in more than
597 50% of AS events (16,182 samples left), and then filtered the AS events using the
598 following criteria: (1) percentage of the missing PSI values in < 5% samples; (2)
599 averaged total reads counts in spliced junctions (i.e., spliced in counts + spliced out
600 counts) > 10; (3) non-constant PSI values across samples; (4) max(PSI) – min(PSI) >
601 0.05; (5) standard deviation > 0.01; (6) averaged PSI value in the range from 0.05 to
602 0.95; (7) TPM of the spliced genes >1. We focused on the AS events in protein-
603 coding genes. In different tissues, the average numbers of genes and AS events for
604 downstream analysis were 12,144 and 13, 939, respectively.

605 Even after removing the AS events with too many missing values, the remaining
606 AS data were still too sparse in some samples for a reliable statistical analysis. On
607 average there are about 0.17% AS events missing from the RNA-seq samples in
608 different tissues. For a single RNA-seq sample in certain tissues, there were up to
609 15% AS events still missing (i.e., below the detecting limit). Several imputation
610 techniques (reference-based and reference-free) have been widely used and easily
611 implemented in multi-omics datasets [70], and we used the k-nearest neighbors (KNN)
612 algorithm for missing PSI value imputation in this analysis.

613

614 **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_i 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.

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 approaches [40]. Linear regression models have been widely used for transcriptomic and genomic differential analysis in recent studies [22,31,34,71], 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:

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

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

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)[72]. 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.

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

Here, Y_{ij} indicates the PSI or TPM of the gene/AS event i in sample j, and the $AveDis(Y_j)$ 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 | Δ 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 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 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 permutated 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:

$$N_i \sim \mu + \alpha Size_i + \beta Detected_i + \varepsilon \qquad (6)$$

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:

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

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 (βM, βF, with cutoffs of p<0.05; GE fold change>1.5 or AS |ΔPSI|>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 *msigdbr* [73], *clusterProfiler* [74], and *DOSE* [75], 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 FastQC [76] (version 0.11.9) and Trim Galore (version 0.6.6). Bowtie2 [77] (version 2.4.2) was used to remove reads mapping to ribosomal RNA. Then, clean reads were mapped to human reference (GRCh38) using STAR [78] (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 Paean [69].

To identify AD-related genes/AS events, we fitted a linear regression model with AD diagnosis (categorical outcomes) and surrogate variables calculated by SVA [72]. 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* [79].

To 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* [80]. 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. S7D). 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 [69]. Differential analysis between control and shRNA groups was performed by a linear regression model:

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

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 [81] to predict the specific binding scores of RBPs within the upstream to downstream region adjacent to the alternative SS. BEDTools getfasta [82] 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

824 p-value < 0.05 were reserved. Taken together, we required the regulatory connections
825 between the RBPs and AS events that pass all the individual criteria as judged by all
826 three types of experiments with different techniques. The networks between age-
827 associated SFs and sBASEs were constructed and figured by Cytoscape [83].
828
829 **Regulation of hormone receptors to the target genes**
830 To examine the estrogen regulation using the RNA-seq dataset, we downloaded
831 the RNA-seq datasets from MCF-7 cell line stimulated by estrogen receptor agonist
832 (propyl pyrazole triol, PPT) in three biological replicates (GEO ID: GSE203058) [45].
833 Additional datasets (ID: GSE89888) are downloaded to confirm our findings in
834 engineered MCF-7 cell lines [46] with wildtype and two constitutively active mutants
835 (Y537S, and D538G) of ESR1. These cells were treated with 1nM estradiol (E2) or
836 vehicle control (veh) for 24 hours before collecting the total RNAs and subjected to
837 transcriptome profiling by RNA-seq. We analyzed these data using the same pipelines
838 in earlier Alzheimer's disease datasets (Fig. 4). The GE changes under estrogen
839 treatments were evaluated in a linear regression model using the p-values of the
840 coefficients deviated from zero. To confirm the expression changes of splicing factors
841 in other cells or tissues, we downloaded the FPKM table from GSE86609 that
842 analyzed the ER knockout mice with estrogen treatments in mouse brain arcuate
843 nucleus (ARCs) [47].
844 To further identify the regulation of the estrogen and androgen using ChIP-seq
845 dataset (Fig. 5E and Fig. 6H), we evaluated the binding scores of ESR1 and AR to the
846 promoter regions of the target genes (i.e., +/- 1kb of the transcription start sites, TSS).
847 The binding score of each target gene (i.e., -10*log10[MACS Q-value] from ChIP-seq
848 analysis) was calculated by the ChIP-Atlas [84], an integrative database of public ChIP-
849 seq data. The ratios in Fig. 5E were calculated by the binding scores of each SFs
850 divided by the median binding score of age-associated SFs common in both sexes.
851 Additionally, the ESR1 ChIP-seq peak files in WT mice treated by estrogen were also
852 downloaded from the Gene Expression Omnibus (ID: GSE3645) [48]. The peaks were
853 annotated using ChIPSeeker [85].
854
855 **Time series and breakpoint analysis**
856 Autoregressive Integrated Moving Average (ARIMA) model is one of the linear
857 models for forecasting chronological trends. For non-stationary data, *auto.arima*()
858 function from R package *forecast* conducts a stepwise procedure to search for the best
859 model using the smallest Akaike Information Criterion (AIC). This procedure is a
860 variation of the Hyndman-Khandakar algorithm [49].
861 We first corrected the TPM/PSI values for each gene/AS event by removing the
862 unknown confounding factors except for the sex and age effect *via* R package *limma*,
863 and averaged the values of the samples with the same sex and age. Then, we applied
864 ARIMA models to figure out genes with potential chronological trends during aging
865 in each sex. In brief, the input data was first transformed into z-ordered objects (R
866 package *zoo*) and the best model was stepwise optimized by AIC (*auto.arima* function
867 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 -log10 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 [83]. 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.

References

1 Westergaard, D., Moseley, P., Sørup, F., Baldi, P. & Brunak, S. Population-wide analysis of differences in disease progression patterns in men and women. *Nature Communications* **10** (2019).

2 Khramtsova, E. A., Davis, L. K. & Stranger, B. E. The role of sex in the genomics of human complex traits. *Nature Reviews Genetics* **20**, 173-190 (2019).

3 Regitz-Zagrosek, V. Sex and gender differences in health: Science & Society Series on Sex and Science. *EMBO reports* **13**, 596-603 (2012).

4 Ferretti, M. T. *et al.* Sex differences in Alzheimer disease—the gateway to precision medicine. *Nature Reviews Neurology* **14**, 457-469 (2018).

5 Jazin, E. & Cahill, L. Sex differences in molecular neuroscience: from fruit flies to humans. *Nature Reviews Neuroscience* **11**, 9-17 (2010).

6 Colafella, K. M. M. & Denton, K. M. Sex-specific differences in hypertension and associated cardiovascular disease. *Nature Reviews Nephrology* **14**, 185-201 (2018).

7 Klein, S. L. & Flanagan, K. L. Sex differences in immune responses. *Nature Reviews Immunology* **16**, 626-638 (2016).

8 Li, C. H., Haider, S., Shiah, Y.-J., Thai, K. & Boutros, P. C. Sex differences in cancer driver genes and biomarkers. *Cancer research* **78**, 5527-5537 (2018).

9 Park, M. D. Sex differences in immune responses in COVID-19. *Nature Reviews Immunology* **20**, 461-461 (2020).

10 Takahashi, T. *et al.* Sex differences in immune responses that underlie COVID-19 disease outcomes. *Nature* **588**, 315-320 (2020).

11 Sampathkumar, N. K. *et al.* Widespread sex dimorphism in aging and age-related diseases. *Human genetics* **139**, 333-356 (2020).

12 Hägg, S. & Jylhävä, J. Sex differences in biological aging with a focus on human studies. *Elife* **10**, e63425 (2021).

13 Austad, S. N. & Fischer, K. E. Sex differences in lifespan. *Cell metabolism* **23**, 1022-1033 (2016).

14 Davis, E. J. *et al.* Sex-specific association of the X chromosome with cognitive change and tau pathology in aging and Alzheimer disease. *JAMA neurology* **78**, 1249-1254 (2021).

15 Ferretti, M. T. & Santuccione Chadha, A. The missing x factor in alzheimer disease. *Nature Reviews Neurology* **17**, 727-728 (2021).

16 Márquez, E. J. *et al.* Sexual-dimorphism in human immune system aging. *Nature communications* **11**, 1-17 (2020).

17 Berchtold, N. C. *et al.* Gene expression changes in the course of normal brain aging are sexually dimorphic. *Proceedings of the National Academy of Sciences* **105**, 15605-15610 (2008).

18 Zhou, X. *et al.* Molecular differences in brain regional vulnerability to aging between males and females. *Frontiers in Aging Neuroscience* **15**, 1153251 (2023).

19 Huang, Z. *et al.* 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** (2021).

20 Han, Y. *et al.* Proteogenomics reveals sex-biased aging genes and coordinated splicing in cardiac aging. *American Journal of Physiology-Heart and Circulatory Physiology* **323**,

H538-H558 (2022).
21 Consortium, G. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* **369**, 1318-1330 (2020).
22 Oliva, M. *et al.* The impact of sex on gene expression across human tissues. *Science* **369** (2020).
23 Khodursky, S. *et al.* Sex differences in interindividual gene expression variability across human tissues. *PNAS nexus* **1**, pgac243 (2022).
24 Zeng, L. *et al.* Transcriptome analysis reveals the difference between "healthy" and "common" aging and their connection with age-related diseases. *Aging cell* **19**, e13121 (2020).
25 Jia, K., Cui, C., Gao, Y., Zhou, Y. & Cui, Q. An analysis of aging-related genes derived from the genotype-tissue expression project (GTEx). *Cell death discovery* **4**, 1-14 (2018).
26 Gallego-Paez, L. M. *et al.* Alternative splicing: the pledge, the turn, and the prestige. *Human genetics* **136**, 1015-1042 (2017).
27 Wang, Z. & Burge, C. B. Splicing regulation: from a parts list of regulatory elements to an integrated splicing code. *Rna* **14**, 802-813 (2008).
28 Baralle, F. E. & Giudice, J. Alternative splicing as a regulator of development and tissue identity. *Nature reviews Molecular cell biology* **18**, 437-451 (2017).
29 Wang, Y., Bao, Y., Zhang, S. & Wang, Z. Splicing dysregulation in cancer: from mechanistic understanding to a new class of therapeutic targets. *Science China Life Sciences* **63**, 469-484 (2020).
30 Karlebach, G. *et al.* The impact of biological sex on alternative splicing. *BioRxiv*, 490904 (2020).
31 Wang, K. *et al.* Comprehensive map of age-associated splicing changes across human tissues and their contributions to age-associated diseases. *Scientific reports* **8**, 1-12 (2018).
32 Bhadra, M., Howell, P., Dutta, S., Heintz, C. & Mair, W. B. Alternative splicing in aging and longevity. *Human genetics* **139**, 357-369 (2020).
33 Angarola, B. L. & Anczuków, O. Splicing alterations in healthy aging and disease. *Wiley Interdisciplinary Reviews: RNA* (2021).
34 Lopes-Ramos, C. M. *et al.* Sex differences in gene expression and regulatory networks across 29 human tissues. *Cell reports* **31**, 107795 (2020).
35 Chan, S., Gomes, A. & Singh, R. S. 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 (2020).
36 Ober, C., Loisel, D. A. & Gilad, Y. Sex-specific genetic architecture of human disease. *Nature Reviews Genetics* **9**, 911-922 (2008).
37 Melé, M. *et al.* The human transcriptome across tissues and individuals. *Science* **348**, 660-665 (2015). https://doi.org/10.1126/science.aaa0355
38 Seidler, R. D. *et al.* Motor control and aging: links to age-related brain structural, functional, and biochemical effects. *Neuroscience & Biobehavioral Reviews* **34**, 721-733 (2010).
39 Balliu, B. *et al.* Genetic regulation of gene expression and splicing during a 10-year period of human aging. *Genome biology* **20**, 1-16 (2019).
40 Li, Y., Ge, X., Peng, F., Li, W. & Li, J. J. Exaggerated false positives by popular differential expression methods when analyzing human population samples. *Genome biology* **23**, 1-

13 (2022).

41 Xiong, J. *et al.* FSH blockade improves cognition in mice with Alzheimer's disease. *Nature*, 1-7 (2022).

42 Matera, A. G. & Wang, Z. A day in the life of the spliceosome. *Nature reviews Molecular cell biology* **15**, 108-121 (2014).

43 Bergles, D. E. & Richardson, W. D. Oligodendrocyte development and plasticity. *Cold Spring Harbor perspectives in biology* **8**, a020453 (2016).

44 Horstman, A. M., Dillon, E. L., Urban, R. J. & Sheffield-Moore, M. The Role of Androgens and Estrogens on Healthy Aging and Longevity. *The Journals of Gerontology: Series A* **67**, 1140-1152 (2012). https://doi.org/10.1093/gerona/gls068

45 Cao, L.-B. *et al.* Estrogen receptor α-mediated signaling inhibits type I interferon response to promote breast cancer. *Journal of Molecular Cell Biology*, mjad047 (2023).

46 Bahreini, A. *et al.* Mutation site and context dependent effects of ESR1 mutation in genome-edited breast cancer cell models. *Breast Cancer Research* **19**, 1-10 (2017).

47 Yang, J. A., Stires, H., Belden, W. J. & Roepke, T. A. The Arcuate Estrogen-Regulated Transcriptome: Estrogen Response Element–Dependent and–Independent Signaling of ER α in Female Mice. *Endocrinology* **158**, 612-626 (2017).

48 Hewitt, S. C. *et al.* Novel DNA motif binding activity observed in vivo with an estrogen receptor α mutant mouse. *Molecular Endocrinology* **28**, 899-911 (2014).

49 Brockwell, P. J., Brockwell, P. J., Davis, R. A. & Davis, R. A. *Introduction to time series and forecasting*. (Springer, 2016).

50 Durieux, J., Wolff, S. & Dillin, A. The cell-non-autonomous nature of electron transport chain-mediated longevity. *Cell* **144**, 79-91 (2011).

51 Djebali, S. *et al.* Landscape of transcription in human cells. *Nature* **489**, p.101-108 (2012).

52 Wang, E. T. *et al.* Alternative isoform regulation in human tissue transcriptomes. *Nature* **456**, 470-476 (2008).

53 Roberts, R. *et al.* The incidence of MCI differs by subtype and is higher in men: the Mayo Clinic Study of Aging. *Neurology* **78**, 342-351 (2012).

54 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).

55 Carithers, L. J. *et al.* A novel approach to high-quality postmortem tissue procurement: the GTEx project. *Biopreservation and biobanking* **13**, 311-319 (2015).

56 Nygaard, V., Rødland, E. A. & Hovig, E. Methods that remove batch effects while retaining group differences may lead to exaggerated confidence in downstream analyses. *Biostatistics* **17**, 29-39 (2016).

57 Van Nostrand, E. L. *et al.* A large-scale binding and functional map of human RNA-binding proteins. *Nature* **583**, 711-719 (2020).

58 Ramanujan, S. A., Cravens, E. N., Krishfield, S. M., Kyttaris, V. C. & Moulton, V. R. 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 (2021).

59 Yoshida, M. & Apte, R. S. NAD+ boosting brings tears to aging eyes. *Nature Aging*, 1-3 (2022).

60 Xiong, X. *et al.* Genetic drivers of m6A methylation in human brain, lung, heart and muscle. *Nature Genetics* **53**, 1156-1165 (2021).
61 Jylhävä, J., Pedersen, N. L. & Hägg, S. Biological age predictors. *EBioMedicine* **21**, 29-36 (2017).
62 Horvath, S. & Raj, K. DNA methylation-based biomarkers and the epigenetic clock theory of ageing. *Nature Reviews Genetics* **19**, 371-384 (2018).
63 Galkin, F., Mamoshina, P., Kochetov, K., Sidorenko, D. & Zhavoronkov, A. DeepMAge: A methylation aging clock developed with deep learning. *Aging and disease* **12**, 1252 (2021).
64 Meyer, D. H. & Schumacher, B. BiT age: A transcriptome-based aging clock near the theoretical limit of accuracy. *Aging Cell* (2021).
65 Sayed, N. *et al.* An inflammatory aging clock (iAge) based on deep learning tracks multimorbidity, immunosenescence, frailty and cardiovascular aging. *Nature Aging* **1**, 598-615 (2021).
66 Galkin, F. *et al.* Human gut microbiome aging clock based on taxonomic profiling and deep learning. *Iscience* **23** (2020).
67 Li, J. *et al.* Determining a multimodal aging clock in a cohort of Chinese women. *Med* (2023). https://doi.org/10.1016/j.medj.2023.06.010
68 Hannum, G. *et al.* Genome-wide methylation profiles reveal quantitative views of human aging rates. *Molecular cell* **49**, 359-367 (2013).
69 Li, J. *et al.* in *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM).* 676-679 (IEEE).
70 Song, M. *et al.* A review of integrative imputation for multi-omics datasets. *Frontiers in genetics* **11**, 570255 (2020).
71 Zhang, Y. *et al.* Regional variation of splicing QTLs in human brain. *The American Journal of Human Genetics* **107**, 196-210 (2020).
72 Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E. & Storey, J. D. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* **28**, 882-883 (2012).
73 Dolgalev, I. msigdbr: MSigDB gene sets for multiple organisms in a tidy data format. *R package version* **7** (2020).
74 Yu, G., Wang, L.-G., Han, Y. & He, Q.-Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics: a journal of integrative biology* **16**, 284-287 (2012).
75 Yu, G., Wang, L.-G., Yan, G.-R. & He, Q.-Y. DOSE: an R/Bioconductor package for disease ontology semantic and enrichment analysis. *Bioinformatics* **31**, 608-609 (2015).
76 Andrews, S. (Babraham Bioinformatics, Babraham Institute, Cambridge, United Kingdom, 2010).
77 Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nature methods* **9**, 357-359 (2012).
78 Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21 (2013).
79 Diedenhofen, B. & Musch, J. cocor: A comprehensive solution for the statistical comparison of correlations. *PloS one* **10**, e0121945 (2015).
80 Kuhn, M. & Johnson, K. *Applied predictive modeling*. Vol. 26 (Springer, 2013).
81 Alipanahi, B., Delong, A., Weirauch, M. T. & Frey, B. J. Predicting the sequence specificities

of DNA-and RNA-binding proteins by deep learning. *Nature biotechnology* **33**, 831-838 (2015).

82 Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842 (2010).

83 Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome research* **13**, 2498-2504 (2003).

84 Oki, S. *et al.* Ch IP-Atlas: a data-mining suite powered by full integration of public Ch IP-seq data. *EMBO reports* **19**, e46255 (2018).

85 Yu, G., Wang, L.-G. & He, Q.-Y. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics* **31**, 2382-2383 (2015).

Acknowledgments

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.

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

Competing interests:

The authors declare that they have no competing interests.

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.

Figures and Tables

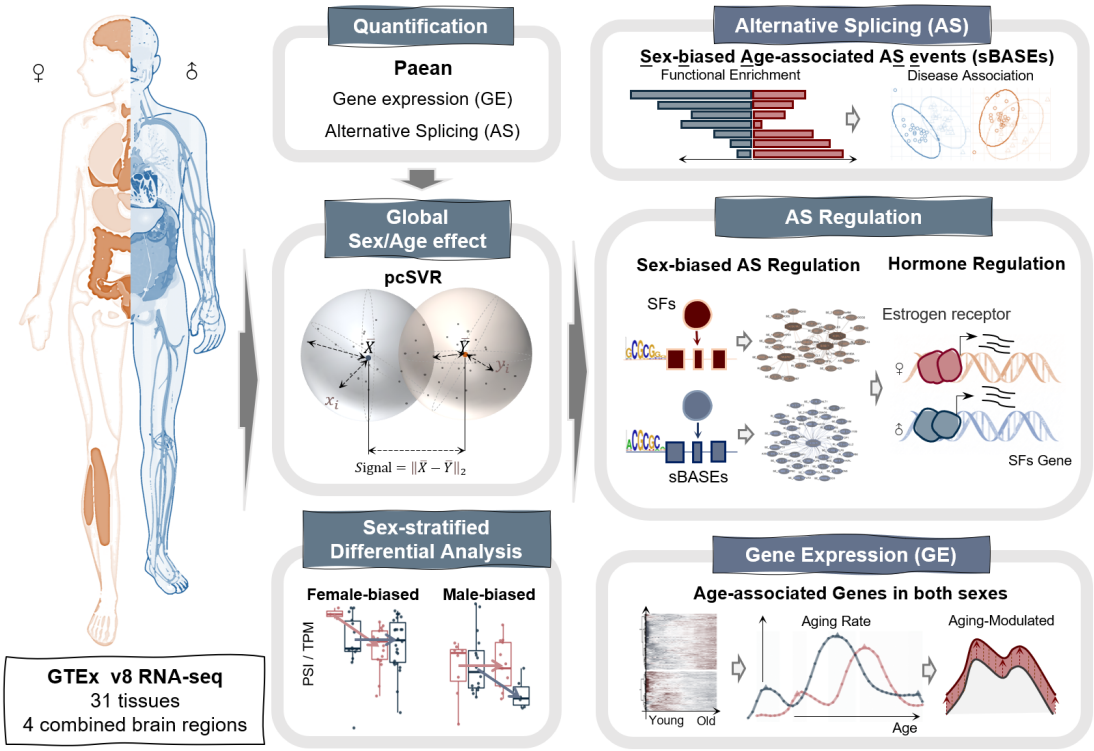


Fig. 1 Schematic Overview of our study.
RNA-seq datasets from 54 human tissues (including 13 brain regions) were downloaded from GTEx Portal. GE and AS quantifications were performed by Paean. The individual sex or age and age-by-sex effects on global transcriptomic variation were evaluated by designing 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.

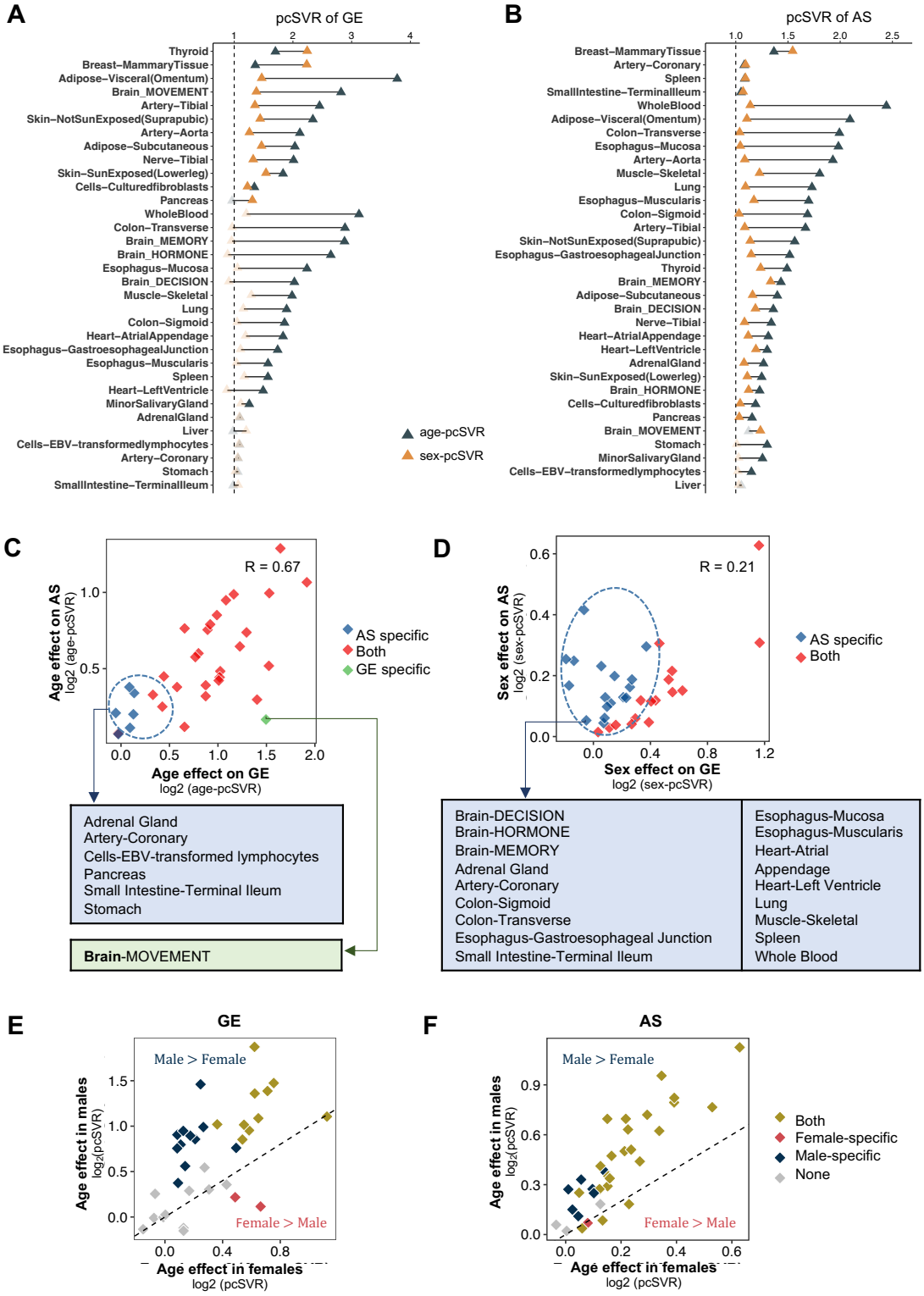


Fig. 2 Individual sex or age effects and combined age-by-sex effects on global transcriptome variation.

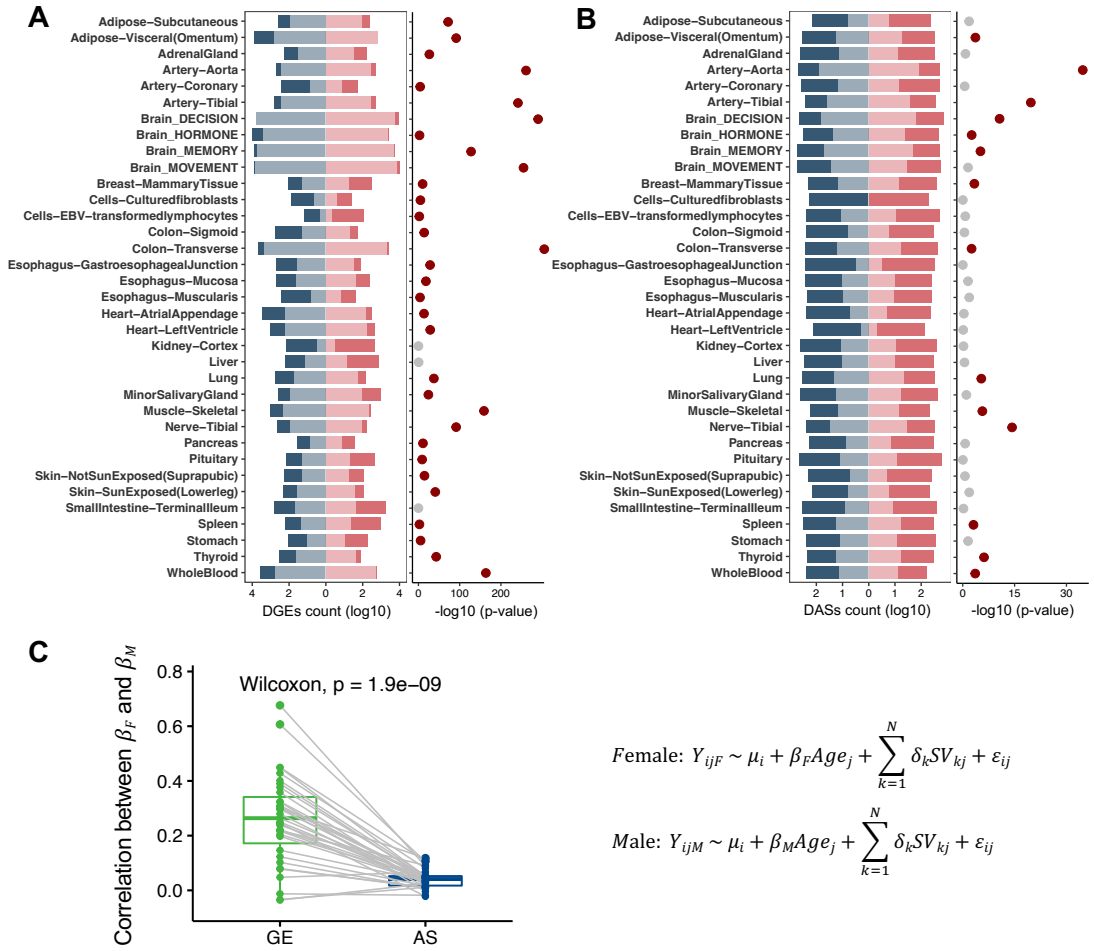
(A-B) pcSVR between different sex or age groups calculated by GE (**a**) and AS (**b**). The yellow triangles show the significant sex-pcSVR calculated between females and males, while black triangles show the significant age-pcSVR calculated between

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 log2 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 log2 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.



$$\text{Female: } Y_{ijF} \sim \mu_i + \beta_F Age_j + \sum_{k=1}^{N} \delta_k SV_{kj} + \varepsilon_{ij}$$

$$\text{Male: } Y_{ijM} \sim \mu_i + \beta_M Age_j + \sum_{k=1}^{N} \delta_k SV_{kj} + \varepsilon_{ij}$$

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.

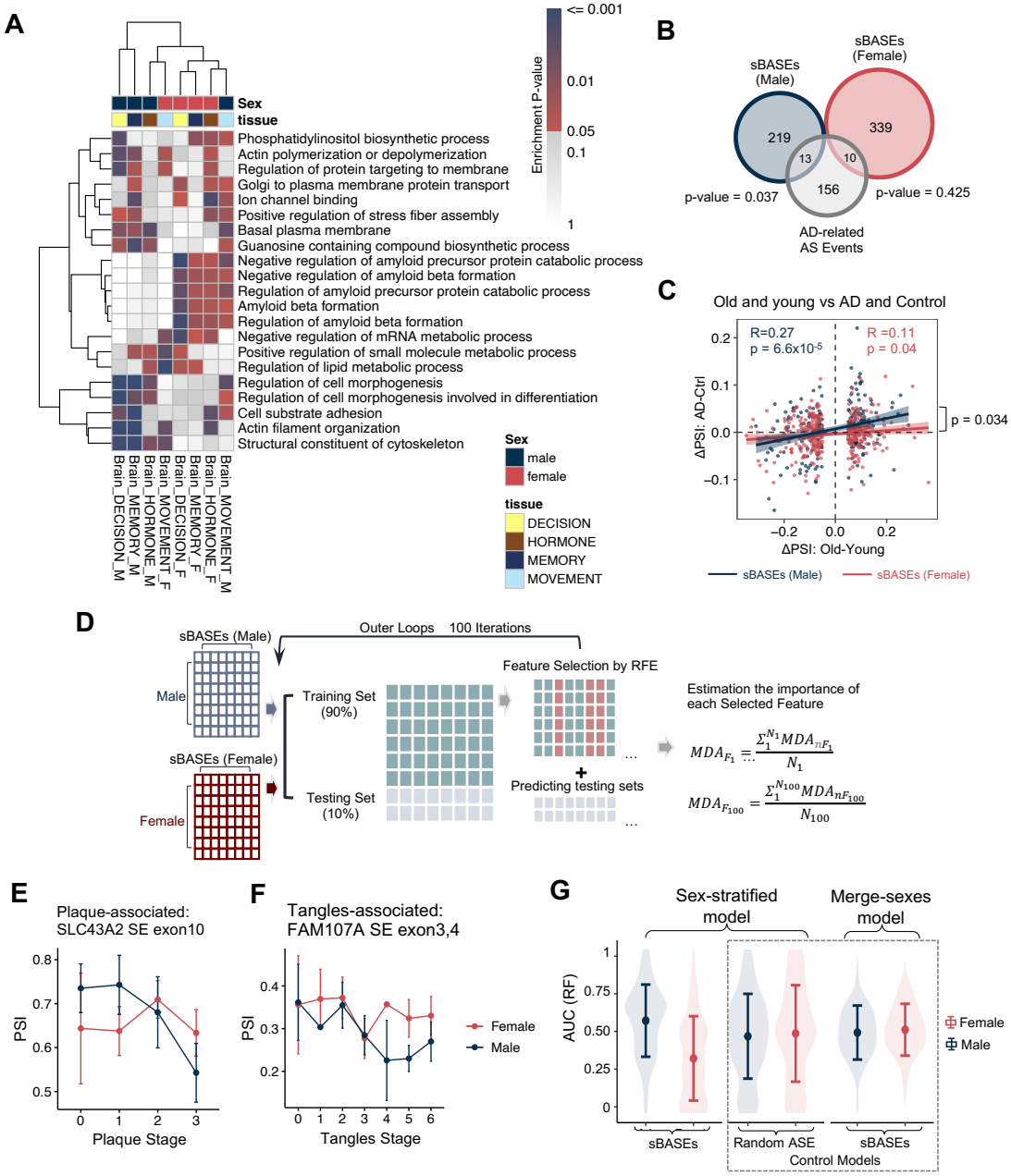


Fig. 4 Male-biased associations between the AS changes during aging and Alzheimer's disease.

(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 -log10 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 PSI_{Old}-PSI_{Young}, while the Y-axis indicates PSI_{AD}-$PSI_{Control}$. 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 ± 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 ± s.d. (standard deviation) of AUC across 100 iterations is shown in the boxplot.

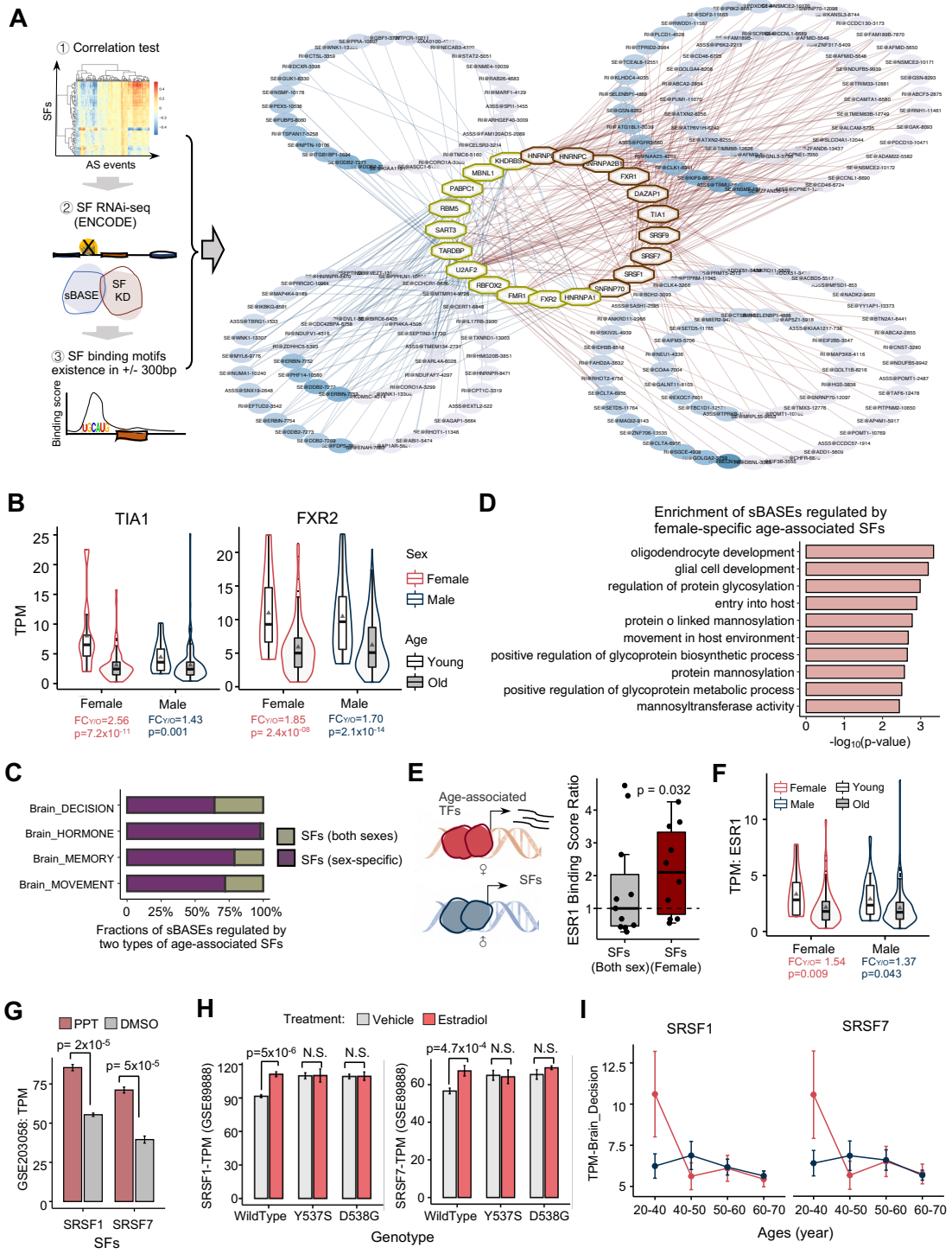


Fig. 5 Sex-dimorphic AS regulation during aging in decision-related brain region.

(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 -log10 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 ± 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, as plotted by mean ± s.e. (standard error).

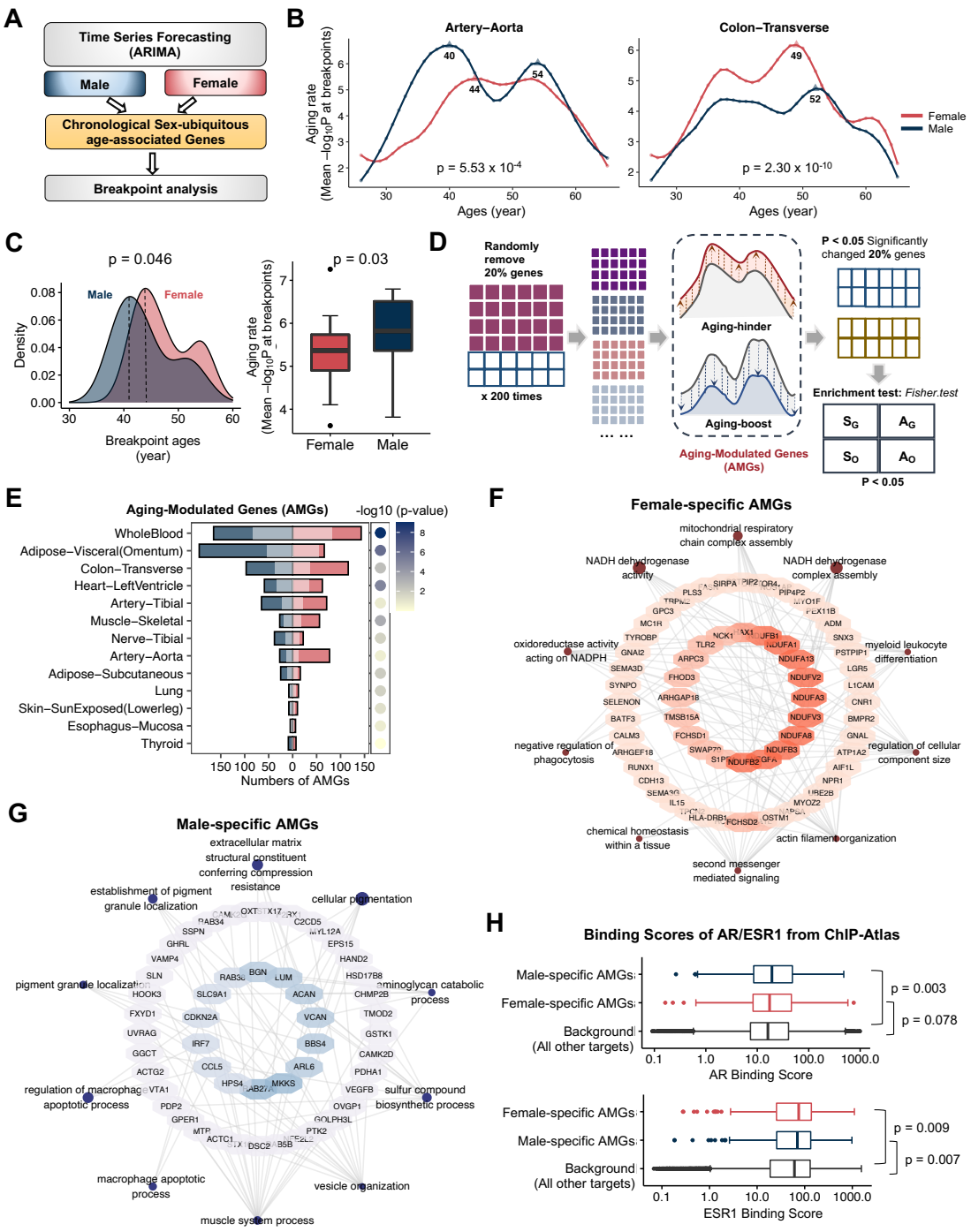


Fig. 6 Sex-dimorphic aging rate of GE during aging process.

(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.