Pan-tissue Transcriptome Analysis Reveals Sex-dimorphic Human Aging

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

[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] CAS Center for Excellence in Molecular Cell Science, University of Chinese Academy of Sciences, Chinese Academy of Sciences, Shanghai, 200031, 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 sex-biased age-associated AS events have a stronger association with Alzheimer's disease in males, and are regulated by several sex-biased splicing factors that are controlled by sex hormones. Breakpoint analysis showed sex-dimorphic aging rates, with males having a larger and earlier transcriptome change, which were significantly associated with decline of sex hormones. Collectively, this study uncovered an essential role of sex during aging at the molecular and multi-tissue levels, providing new insight into sex-dimorphic regulatory patterns.

Keywords: Genotype-tissue Expression (GTEx); Sex-dimorphism; Aging; Alternative Splicing (AS); AS regulatory network; Time series

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. New 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, have been studied using the blood of healthy donors [16]. 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 and disease susceptibility [17]. 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 [18,19]. Previous studies have uncovered many sex-biased genes or expression quantitative trait loci (eQTLs)[19] 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 [20,21]. 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 [22]. AS is tightly regulated in various tissues or development stages by multiple *cis*-acting elements and *trans*-acting factors [23,24]. Dysregulation of AS is closely related to many age-related diseases such as cancer [25]. 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 [24]. Recent studies reported that AS regulation is closely associated with sex [26] and age [27,28]; 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 studied 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 [18] 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 (Supplementary Table 1). We performed the principal component analysis (PCA) on GE and AS data, and observed the sex and age differences in several tissues (Supplementary Fig. 1). Based on a previous study [29], we designed a new 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 (Supplementary Fig. 2a, see 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. 2a, b). Due to a large variation of menopause ages among different individuals and a continuous decline of sexual steroid levels between age 40 and 60 [30,31], we grouped the samples into young (age<40) *vs.* old (age>60) with an age gap instead of a specific cutoff age 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 [29], 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 judged by both GE and AS (Fig. 2a, b, Supplementary Table 2). 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 2a and 2b). Similar results were found even when removing the genes and AS events encoded by sex-chromosomes (Supplementary Fig. 2b, c), selecting a range of PC cutoffs to capture different global variance (Supplementary Fig. 2d, e), or using other approaches for p-value estimation (e.g., boot-strapping with replacement, see Methods, Supplementary Fig. 2f, g).

We further compared the age and sex effects calculated by GE *vs.* AS (Supplementary Table 2), 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. 2a, 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. 2e, 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. 2e, f). These results suggested that age and sex contributed significantly to transcriptome variation in various tissues with certain degrees of heterogeneity for different data types.

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

We next identified specific genes or AS events responsible for sex or age effect on transcriptomic variations *via* a linear regression model containing an age-by-sex coefficient term (see Methods, Equations 4 and 5). In each tissue, we identified the sex- or age-differential genes and AS events (Supplementary Fig. 3a, b). Consistent with previous reports [29,32,33], the breast tissue contains the largest amount of sex-differentially expressed genes, while the age-biased genes were widely discovered in brain regions and artery-related tissues. Interestingly, in most tissues, there were no significant overlaps between age-differential genes and AS events (Supplementary Fig. 3a, hypergeometric test p-value < 0.01). In contrast, we found significant overlaps between sex-differential genes and AS events in most tissues (Supplementary Fig. 3b).

As a control, we constructed a separate regression model to evaluate the contribution of sampling biases, including sample sizes and the total number of genes or AS events detected in each tissue (Equation 6). We found that the sample size did not significantly affect the number of identified sex-differential or age-differential genes/AS events (Supplementary Fig. 3c, d), suggesting our analyses are not affected by sampling biases. Furthermore, using the coefficient of the age-by-sex term in this model allowed us to identify thousands of genes and AS events affected by the functional interactions between sex and age (i.e., the sex effects on GE/AS depend on the age, *or vice versa*) (Supplementary Fig. 3e), suggesting extensive sex-dimorphic changes of transcriptome during aging.

A recent study reported the environmental effect on human genetic background during aging [34], which should be considered when analyzing large datasets from different populations. In the linear regression model, we included the surrogate

166 variables (SV) as confounders to control for the effect of genetic backgrounds
167 (Equations 4 and 5). Using the differential GE analysis in whole blood as an example
168 (other tissues have the same trend), we evaluated the correlation between SVs with
169 donors' ethnicity using Point Biserial Correlation, and found that many SVs showed
170 significant correlations with the donors' ethnical background (Supplementary Fig. 4a),
171 suggesting that these SVs could effectively reflect the variables of genetic background.
172 Additionally, we calculated the correlation between SVs with the top5 principal
173 components as judged by whole genome sequencing (WGS) data (Supplementary Fig.
174 4b), and found roughly the same set of SVs also showed significant correlation with
175 PC1. Therefore, the surrogate variables in the linear regression models could control
176 the cofounders of human genetic regulation.

177 To further evaluate the robustness of our differential analyses, we performed
178 detailed assessments of the linear regression models and the selection of GE/AS cutoffs
179 using the decision-related brain region as an example. We calculated the false discovery
180 rates (FDR) using the permutation analysis in the linear regression model (see Methods).
181 Briefly, we generated the permuted data 100 times by randomly shuffling the group
182 labels (sex and age) with the same sample sizes as the original labels. The distribution
183 of the p-values of sex/age-differential genes/AS events identified based on the original
184 and each shuffled phenotype data were evaluated (Supplementary Fig. 4c, 4e). The FDR
185 is defined as the fraction of the original differential genes/AS events identified also
186 from permutated data. We found low FDRs of sex/age-differential genes/AS events
187 using the linear regression model (Supplementary Fig. 4d, 4f), which is much lower
188 than the FDR of DESeq2 and edgeR [35]. As a result, this approach could justifiably
189 control the false discovery for differential analysis.

190 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 (Supplementary Fig. 5a-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 (Supplementary Fig. 5c), which is consistent with our earlier findings (Supplementary Fig. 3a). 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 (Supplementary Table 3). 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 (Supplementary Fig. 5d). 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 neurodegenerative diseases

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 (Supplementary Fig. 6a, Supplementary Table 4). 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 (Supplementary Fig. 6b). 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 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., $PSI_{Old} - PSI_{Young}$) (P=0.001, R=0.22), but such correlation was not discovered in a separate analysis in females using the female-specific sBASEs (P=0.103, R=0.08; Fig. 4c). 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)

241 (Supplementary Fig. 6c), supporting a recent finding that females are more vulnerable
242 to AD during aging due to the post-menopause activation of C/EBPβ–AEP/δ-secretase
243 pathway by sex hormone changes [36].

244 In addition, we performed similar analyses using an independent dataset from
245 ROSMAP containing 271 filtered RNA-seq samples from dorsolateral prefrontal cortex
246 (DLPFC) with matched age distribution between AD and control samples. As expected,
247 we found that the AD-associated AS changes of sBASEs in males showed a stronger
248 positive correlation with the age-associated changes ($R=0.469$, $p < 2.2\times10^{-16}$) than such
249 correlation across the sBASEs in females ($R=0.364$, $p < 2.2\times10^{-16}$) (Supplementary Fig
250 6d), which is largely consistent with the earlier finding using a smaller dataset (Fig. 4c).
251 This male-biased association indicated that the AS changes during aging could
252 contribute more to the AD in males than females, suggesting additional molecular
253 complexity in the sex effect of AD.

254 To further examine the contribution of individual sBASE to AD risk, we
255 constructed a sex-stratified random forest classifier using a subset of sBASEs for AD
256 prediction. For each sex, we randomly divided the datasets of AD patients and controls
257 into separate training and test sets, and repeated the prediction 100 times (Fig. 4d). For
258 each selected sBASE, we also evaluated its averaged feature importance (i.e., mean
259 decrease accuracy) across 100 iterations and the association with the classic AD
260 neuropathology (such as tangles and plaques), resulting in the identification of a series
261 of key sBASEs significantly associated with AD in a sex-specific manner. For example,
262 the skipping of exon 10 in SLC43A2 negatively correlated with the neuritic plaques
263 stage in males but not in females (Fig. 4e). Meanwhile, skipping of exons 3 and 4 in
264 FAM107A, a gene related to synaptic and cognitive functions, showed a male-specific
265 negative association with neurofibrillary tangle stages (Fig. 4f). In addition, the

266 prediction models based on sBASEs performed significantly better in males than
267 females (Fig. 4g, left panel), suggesting that age-associated AS events could serve as a
268 better predictor for AD risk in male patients. As negative controls, we performed the
269 same analysis to predict male and female patients separately using randomly selected
270 AS events (sex-stratified model; Fig. 4g, middle panel), and predict all patients with
271 male- or female-specific sBASEs (merge-sexes model; Fig. 4g, right panel). We found
272 that both control models showed no differences in males and females, indicating the
273 crucial roles of sBASEs in AD during male aging. Collectively, our analyses of human
274 brains revealed a surprising sex-biased association between RNA splicing and AD,
275 suggesting that regulation of AS may play a critical role in the sex-dimorphisms of age-
276 related diseases such as AD.
277
278 **sBASEs are regulated by splicing factors with sex-dimorphic expression during**
279 **aging.**
280 AS is generally regulated by splicing regulatory *cis*-elements that specifically
281 recruit various *trans*-acting splicing factors (SF) to promote or inhibit splicing of
282 adjacent exons [23,37]. Dysfunction of the SFs has been reported as a hallmark of aging,
283 presumably by regulating the genes in cellular senescence during aging process [28].
284 Consistently, our analyses identified a large number of age-associated SFs in human
285 brain (Supplementary Table 5). To further examine the sex-dimorphic regulation
286 through age-associated SFs, we constructed an SF-RNA regulatory network between
287 age-associated SFs and sBASEs in each sex. This process integrated the correlations
288 between the SFs levels and PSI of sBASEs, the significant changes of sBASEs upon
289 SFs knockdown, and the existence of SFs binding motifs around sBASEs (Fig. 5a, left
290 panel). We required the regulatory connections between the SFs and AS events that pass

291 all these three criteria. Such regulatory networks can be generated in different tissues,
292 and we focused on the brain due to its association with age-related diseases such as
293 cognitive disorders and synaptic diseases.

294 We identified several key age-associated SFs that regulate multiple sBASEs in the
295 decision-related brain region (Fig. 5a, right panel). Interestingly, several SFs showed
296 significant age association only in females (e.g., TIA1 in Fig. 5b, left panel), while
297 others were associated with age in both sexes (e.g., FXR2 in Fig. 5b, right panel). This
298 finding is consistent with our earlier observation that more sex-specific age-associated
299 AS events were identified in decision-related brain region of females than males (Fig.
300 4b). Most of the sBASEs were regulated by sex-specific age-associated SFs rather than
301 the SFs associated with aging in both sexes (Fig. 5c). Functional enrichment analysis
302 showed that these sBASEs regulated by female-specific age-associated SFs were
303 significantly enriched in glial cells and oligodendrocytes development (Fig. 5d), which
304 participated in the myelin generation of aging brain and neurodegenerative diseases in
305 central nervous system [38]. These results indicated that the SFs with sex-dimorphic
306 expression during aging, especially the female-specific changes of SFs, were one of the
307 main factors in regulating sBASEs with important biological functions.

308 Furthermore, we studied the potential mechanism that controls the female-specific
309 changes of SFs during aging in decision-related brain region. It was previously reported
310 that the female sex hormone estrogen consistently decreases during lifespan [31,39]; thus
311 we speculated that estrogen may play roles in controlling transcription of SFs during
312 aging through activation of nuclear receptors. To this end, we analyzed the public ChIP-
313 Atlas data to examine the binding of canonical estrogen receptor ESR1 to the promoter
314 region. We found that the ESR1 binding ratios on the female-specific age-associated
315 SFs were significantly higher than the median value of the age-associated SFs common

316 in both sexes (Fig. 5e; p = 0.032), suggesting that ESR1 was more likely to bind the
317 promoters of the female-specific age-associated SFs. In addition, the age-associated
318 decline of ESR1 was more substantial in females than males (Fig. 5f), and some the
319 female-specific age-associated SFs (e.g., SRSF1 and SRSF7) could be stimulated by
320 propyl pyrazole triol, an agonist of ERα (Fig. 5g). These observations were further
321 confirmed by an independent experiment where the wildtype and two constitutively
322 active mutants (Y537S, and D538G) of ESR1 were engineered into MCF-7 cells using
323 CRISPR-Cas technology [40]. The resulting cells were treated by estradiol or vehicle
324 controls and subjected to transcriptome profiling by RNA-seq [40], and the data were
325 reanalyzed using the same pipeline. We found that the basal levels of SRSF1 and
326 SRSF7 were significantly higher in the cells with constitutively active ESR1 (Fig. 5h,
327 comparing samples treated with vehicle controls), and SRSF1 and SRSF7 were further
328 induced by the estradiol treatment only in cells with wildtype ESR1 but not in cells with
329 mutated ESR1 (Fig. 5h, compare the samples treated with estradiol *vs.* control in
330 different cells), indicating that the activation of ESR1 pathway can promote expression
331 of SRSF1 and SRSF7. As expected, the expression levels of SRSF1 and SRSF7 showed
332 an obvious decrease during female aging (Fig. 5i). Collectively these data suggested
333 that the female-specific SFs (e.g., SRSF1 and SRSF7) are more sensitive to the estrogen
334 changes *via* ESR1-mediated pathway, leading to extended sex-biased AS regulations
335 during aging.
336
337 **The aging rates of GE show sex-dimorphism during aging across multiple tissues**
338 In the earlier results (Fig. 3a), we found a large proportion of age-associated genes
339 are common in both sexes. Therefore, we asked whether the levels of GE change at the
340 same pace during aging in females and males. The age-associated genes ubiquitously

341 changed in both sexes were analyzed for consistency. We first used the Autoregressive
342 Integrated Moving Average model [41] to capture the age-associated genes with
343 significant chronological changes beyond random fluctuations, which are the main
344 contributor to transcriptome dynamics during aging. We further quantified the GE
345 changes between adjacent age windows using breakpoint analysis [16] (Fig. 6a, see
346 Methods). Such analyses were conducted across all tissues in both sexes
347 (Supplementary Fig. 7a), with two tissues shown in Fig. 6b. A higher value on the Y-
348 axis (mean -log P) indicated a larger GE change at the given age, and thus the peaks
349 were regarded as aging "breakpoints" that represent the key time point of transcriptome
350 aging. Compared to females, males showed significantly larger GE changes during
351 aging in most tissues (Supplementary Fig. 7a), except transverse colon, fibroblast cells,
352 and whole blood (Supplementary Fig. 7b). The number of breakpoints also showed sex-
353 dimorphism in several tissues. For example, in sun-exposed skin and colon, males
354 showed two main breakpoints around ages 35 and 50, whereas females only have one
355 breakpoint at age ~45 (Supplementary Fig. 7a).

356 Moreover, we found that the major aging breakpoints (i.e., breakpoints with the
357 largest GE changes) occurred earlier in males than females across most tissues, and the
358 magnitudes of changes were also more prominent in males (Fig. 6c), suggesting that
359 males may age faster with a larger age-associated GE change. In addition, we conducted
360 breakpoint analysis with all chronological genes in both sexes and found similar results
361 when comparing the breakpoints in males and females (Supplementary Fig. 7c). We
362 also performed a similar analysis using chronological sex-biased age-associated AS
363 events (sBASEs) and all chronological AS events in females and males separately, and
364 again found an earlier and more obvious breakpoint (Supplementary Fig. 7d-e). In
365 summary, these results uncovered an earlier aging and a higher aging rate in males at

366 molecular levels, again implying that males could be surprisingly more vulnerable to
367 aging.
368
369 **Aging contributors in separate sexes show distinct functions**
370 To further examine how the individual gene affected aging rates, we designed a
371 disturbance analysis to identify aging-disturbance genes (ADGs) (Fig. 6d). Specifically,
372 we randomly removed 20% of chronological genes and repeated the breakpoint analysis
373 with the remaining genes to test if the breakpoints were significantly altered (see
374 Methods). This process was repeated 200 times, and the genes enriched in the 200
375 random samplings (judged by Fisher exact test $p < 0.05$) were defined as ADGs that
376 serve as the key contributors to molecular aging.
377 We identified total 1035 ADGs across different tissues (Supplementary Table 6),
378 around 70% of which (723 in 1035) were sex-specific (Fig. 6e). Functional enrichment
379 analysis suggested that the female-specific ADGs were significantly enriched in cellular
380 respiration and NADH dehydrogenase pathways (Fig. 6f), which is consistent with an
381 earlier report of their roles in human aging and longevity [42]. Such functional enrichment
382 was not found in the male-specific ADGs (Fig. 6g), implying that female aging may be
383 more sensitive to mitochondrial and oxidative functions. Intriguingly, we also found
384 many male-specific ADGs relevant to myeloid leukocyte differentiation, phagocytosis,
385 and macrophage apoptotic processes, implying a sex-differential role of the immune
386 system during aging.
387 To determine the potential mechanism of the sex-dimorphic aging rate, we focused
388 on sex hormones that are critical for transcriptional regulation. We found that the
389 androgen receptor (AR) tends to bind the promoters of male-specific ADGs (P=0.007),
390 suggesting potential regulations by androgen during male aging (Fig. 6h, top panel;

Supplementary Fig. 7f). On the other hand, female-specific ADGs were more likely to be bound and regulated by estrogen receptor ESR1 (Fig. 6h, bottom panel; Supplementary Fig. 7g), suggesting the potential regulation by distinct sex hormone.

We further examined the detailed expression profile of ADGs from the whole blood that is commonly used for health surveillance. The ADGs 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 (Supplementary Fig. 8a, 7b). These genes were functionally enriched in neutrophil and granulocyte migration and RNA splicing (Supplementary Fig. 8c). Moreover, we also examined sex-specific ADGs and found very distinct expression patterns (Supplementary Fig. 8d, e). Interestingly, while most of these genes showed decreased expression at old ages, the expression of some male-specific ADGs gradually increased during aging (Supplementary Fig. 8d, dotted box). Taken together, these results suggested that these age-associated genes disturbing 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 RNA-seq 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 picture 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 cell types and tissues [43,44]. 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, as well as a higher discrepancy between females and males from the combinatorial effects of age-by-sex, highlighting the contribution of AS 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 hormone [36], whereas a higher incidence rate of mild cognitive impairment in males is also reported [45]. At the splicing level, we surprisingly observed a male-biased association between the age-associated AS changes and AD. However, such sex-biased associations could not be detected by analyzing GE, suggesting that analysis at isoform resolution can provide new 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 for 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 the transcriptomic aging happened earlier in males than in females (Fig. 6d), consistent with the reported shorter life expectancy in males [12]. Furthermore, we identified many sex-specific ADGs with different biological functions, such as the macrophage and mitochondria functions (Fig. 6f, g). These genes tend to be regulated by the dominant sex hormones in each sex. The investigation of the aging rate provided insights into the intervention of the aging process or age-related diseases, and to some extent made available for novel therapeutic targets at specific age points. It is worth noting that our analysis of the aging rate in this section does not include the data from brain tissues due to the sample sparsity in selecting parameters.

Our analyses also have some technical limitations, the biggest of which may be

456 caused by the cell type and sample heterogeneity between different donors of GTEx
457 datasets. Although this is an intrinsic problem associated with big data and cannot be
458 completely removed, we used surrogate variables in differential analysis to reduce such
459 heterogeneity. Similarly, the sample numbers for some tissues are small, and thus we
460 either omitted them from aging analyses (like pituitary, bladder, and kidney) or
461 combined similar tissues (like brain regions) for detailed analyses. The publicly
462 available shRNA-seq and CLIP-seq data from literature or ENCODE database also
463 affected our analyses in constructing AS regulatory network, as some of the SFs or cell
464 types are missing.

465 Mechanistically, we found a prominent role of sex hormones in regulating sex
466 dimorphic aging or age-related diseases. Using a range of RNA-seq and ChIP-seq data,
467 we uncovered a new type of female-specific age-associated splicing regulation (Fig.
468 5a), which may partially be achieved through the estrogen-mediated regulation on the
469 expression of female-specific splicing factors (Fig. 5e-5i). We further quantified
470 chronological changes of transcriptome dynamics using breakpoint analysis, and
471 observed the sex-dimorphic aging rates across multiple tissues. The aging breakpoints
472 we identified is largely consistent with the changes of circulating sex hormones during
473 aging. The level of testosterone starts to decrease at the age 35–40 in males, while in
474 females, estrogen level starts to decrease at a later age (i.e., ~50 years old) [31,39].
475 Mechanistically, we identified a series of female-specific and male-specific aging-
476 disturbance genes (ADGs), whose functions are consistent with the reported roles of
477 sex hormones. For example, the estrogen was reported to participate in the antioxidant
478 system *via* NADPH [31,39] , which we found to be an enriched function in female-specific
479 ADGs (Fig. 6f). Conversely, the androgen was reported to affect the inflammation in
480 immune response [31,39], which is an enriched function in the male-specific ADGs (Fig.

481 6g). We further examined the causative relationship between these ADGs and sex
482 hormones using ChIP-Atlas, and found that the sex-specific ADGs were more likely to
483 be regulated by androgen receptor in males or estrogen receptor in females (Fig.6h).
484 However, the regulation of sex-biased aging is probably more complicated. Other
485 factors, including mitochondrial functions, genes encoded by sex chromosomes, non-
486 coding RNAs, and transposable elements, may also contribute to the sex-dimorphism
487 during aging [12,14,46,47]. In addition, it is possible that the genetic variants, including
488 expression QTL (eQTLs) and splicing QTLs (sQTL), can affect GE or AS in a sex-
489 biased or age-associated fashion [19,34,48]. Future studies of these factors will improve our
490 understanding of the intricate regulation of transcriptome dynamics, providing us a
491 broader horizon at a multi-omics scale.

492 Various biological aging clocks had previously been constructed based on different
493 molecular markers, including DNA methylome [49,50], transcriptome [51], circulating
494 immune proteins [52], and human gut microbiome [53]. Such aging clocks generally ignored
495 the difference between females and males. However, our results advocate for a sex-
496 specific age model. In addition, it has been proposed that certain age-related diseases
497 can accelerate aging process, as tumors aged 40% faster than matched normal tissues
498 based on DNA methylome [54]. Therefore, we speculate that constructing separate aging
499 clocks in distinct sexes may help model the progression of age-related diseases
500 accurately.
501

502 **METHODS**

503 **Transcriptome quantifications and data processing**

504 We used the transcriptome data from Genotype-Tissues Expression (GTEx) project
505 version 8.0, which contains 17,382 samples in 54 tissues (including 13 brain regions
506 and two cell lines). We downloaded the RNA-seq bam files and the phenotype
507 information from dbGaP (study accession: phs000424.v8.p2; table accession:
508 pht002743.v8.p2.c1). The tissues with a small number of samples and specific to only
509 a single sex (such as ovary and testis) were removed for a better comparison. In
510 addition, for each donor, we combined 13 brain regions into 4 main functional regions,
511 including hormone- or emotion-related region (amygdala, anterior cingulate cortex, and
512 hypothalamus), movement-related region (cerebellar Hemisphere, cerebellum, and
513 spinal cord), memory-related region (caudate, hippocampus, nucleus accumbens,
514 putamen, and substantia nigra) and decision-related region (cortex, and frontal cortex).
515 As a result, for the N donors that each have 13 brain regions (N×13 sample matrix), we
516 merged the data into 4 functional regions to generate an N×4 sample matrix in the
517 downstream analyses. In total, 16,202 samples in 35 tissues or tissue groups from 948
518 donors passed these filters for further analysis. The numbers of samples in each tissue
519 are shown in Supplementary Table 1.

520 GE and AS quantifications were performed by Paean [55], a parallel computing
521 system on the GPU-CPU platform with high computational efficiency for super-large
522 datasets. We calculated TPM (transcripts per million) for protein-coding genes and PSI
523 (percent of spliced in) values for AS events, including skipped exon (SE), alternative 5′
524 splice site (A5SS), alternative 3′ splice site (A3SS), and retained intron (RI). Genes
525 with low expression (average TPM < 1) were removed. We also filtered the AS events

526 by firstly removing the samples with missing PSI values in more than 50% AS events
527 (16,182 samples left), and then filtered the AS events using the following criteria: (1)
528 percentage of the missing PSI values in < 5% samples; (2) averaged total reads counts
529 in spliced junctions (i.e., spliced in counts + spliced out counts) > 10; (3) non-constant
530 PSI values across samples; (4) max(PSI) – min(PSI) > 0.05; (5) standard deviation >
531 0.01; (6) averaged PSI value in the range from 0.05 to 0.95; (7) TPM of the spliced
532 genes >1. We focused on the AS events in protein-coding genes. In different tissues, the
533 average numbers of genes and AS events for downstream analysis were 12,144 and 13,
534 939, respectively.

535 Even after removing the AS events with too many missing values, the remaining
536 AS data were still too sparse in some samples for a reliable statistical analysis. On
537 average there are about 0.17% AS events missing from the RNA-seq samples in
538 different tissues. For a single RNA-seq sample in certain tissues, there were up to 15%
539 AS events still missing (i.e., below the detecting limit). Several imputation techniques
540 (reference-based and reference-free) have been widely used and easily implemented in
541 multi-omics datasets [56], and we used the k-nearest neighbors (KNN) algorithm for
542 missing PSI value imputation in this analysis.

543

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

545 To evaluate the sex and age effect on human transcriptome, we designed a principal
546 component-based signal-to-variation ratio (pcSVR) based on principal component
547 analysis (PCA) and signal-to-noise ratio (SNR). PCA process could maximize the
548 global transcriptomic variations, and SNR could calculate the distances between
549 sex/age groups divided by the noises within each group (Equations 1-3, female *vs.* male
550 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 numbers 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. Compare 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 [35]. Linear regression models have been widely used for transcriptomic and genomic differential analysis in recent studies [19,27,29,57], 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)[58]. For differential GE analysis, we used the same linear regression

599 model for analyzing differential GE by replacing PSI with TPM in the formula
600 (Equation 5). To deal with the different orders of magnitude between PSI and TPM
601 values, we normalized the PSI/TPM of each gene/AS event by their averaged PSI/TPM
602 values across samples. We defined sex-differential or age-differential genes with the
603 cutoffs of p-value < 0.05 for the α or β coefficients, and foldchange >1.5. Sex-
604 differential or age-differential AS events were defined with the cutoffs of p-value < 0.05
605 for the α or β coefficients, and |ΔPSI|>0.05. The γ significantly larger or smaller than
606 0 indicates significant age-by-sex interaction, which suggests that the values of TPM
607 and PSI are not only determined by sex or age independently but also affected by the
608 age at specific sex or sex at a specific age.

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

617 Permutation analysis was used to address the false discovery issue of this linear
618 regression model in the differential analysis. We first generated the permuted data by
619 randomly shuffling the group labels (sex and age) with the same sample sizes as the
620 original group labels. The differential genes/AS events were identified based on each
621 generated permutated phenotype data. This process was performed 100 times and the
622 FDR is defined as the fraction of the original differential genes/AS events also
623 identified from permutated data.

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 surrogate variables (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.

649 In addition, ΔPSI and Foldchange have also been used for filtering, which are the
650 important thresholds to evaluate the degree of changes. Thus, we further defined the
651 female- or male-specific age-associated genes/AS events using this sex-stratified
652 analysis, and classified these age-associated genes/AS events into sex-biased age-
653 associated genes/ ASs events (i.e., age-associated genes or AS events specific to one
654 sex) or sex-ubiquitous age-associated genes/AS events (i.e., age-associated genes or AS
655 events in both sexes).
656
657 **Functional enrichment and Disease Ontology analysis**
658 Potential functions of sBASEs were expounded based on C5 collections (ontology
659 gene sets) in Molecular Signatures Database (MSigDB) and Disease Ontology. We
660 carried out these processes using Bioconductor R packages *msigdbr* [59], *clusterProfiler*
661 [60], and *DOSE* [61], with the significance threshold p-value <0.05. The GO terms with the
662 significances specific in only one sex per tissue were selected for downstream analysis.
663 In the enrichment heatmap of brain regions (Fig. 4a), we plotted the GO terms with
664 significance in at least 3 brain regions. Because the enrichment heatmap across all
665 tissues are unrealistically large, we plotted the terms within at least 15 tissues and the
666 tissues/sex with significance in at least 9 significant GO terms (Supplementary Fig. 6a).
667 The entire results of the enrichment analysis were shown in Supplementary Table 4.
668
669 **Processing of Alzheimer's disease datasets and construction of AD prediction**
670 **model**
671 The AD samples of human prefrontal cortex were selected from GEO datasets (ID:
672 GSE174367) containing 375 samples of single-nucleus RNA-seq, single-nucleus
673 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 [62] (version 0.11.9) and Trim Galore (version 0.6.6). Bowtie2 [63] (version 2.4.2) was used to remove reads mapping to ribosomal RNA. Then, clean reads were mapped to human reference (GRCh38) using STAR [64] (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 [55].

To identify AD-related genes/AS events, we fitted a linear regression model with AD diagnosis (categorical outcomes) and surrogate variables calculated by SVA [58]. AD-related genes/AS events were defined whose coefficients of the 'diagnosis' term significantly deviated to zero.

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 (Fig. 4d). In each iteration, we performed a 5-fold cross-validation with the feature selection approach called Recursive Feature Elimination (RFE) via R packages *caret* [65]. Model performance was evaluated by the area under the curve (AUC). 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 [55]. 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 [66] to predict the specific binding scores of RBPs within the upstream to downstream region adjacent to the alternative SS. BEDTools getfasta [67] extracted sequences from specific genomic regions in a .bed6 file. We first subdivided the sequences into multiple subsequences (window = 40nt; step =1nt) and calculated the binding scores with the parameter 'average' of each subsequence. To identify whether there were binding signals in the upstream to downstream 300nt region adjacent to the alternative splice sites, we rearranged the binding scores in each region into small bins (width = 20), and tested the median differences between the bins with maximum and minimum averaged binding scores by Wilcoxon rank-sum test. The AS events whose maximum binding scores > 1 and p-value < 0.05 were reserved. Taken together, we required the regulatory connections between the RBPs and AS events that pass all the individual criteria as judged by all three types of experiments with different techniques. The networks between age-associated SFs and sBASEs were constructed and figured by Cytoscape [68].

Regulation of hormone receptors to the target genes

To examine the estrogen regulation using RNA-seq dataset, we downloaded the RNA-seq datasets from MCF-7 cell line stimulated by estrogen receptor agonist (propyl pyrazole triol, PPT) in three biological replicates (GEO ID: GSE203058). Additional datasets (ID: GSE89888) are downloaded to confirm our findings in engineered MCF-7 cell lines [40] with wildtype and two constitutively active mutants (Y537S, and D538G)

749 of ESR1. These cells were treated with 1nM estradiol (E2) or vehicle control (veh) for
750 24 hours before collecting the total RNAs and subjected to transcriptome profiling by
751 RNA-seq. We analyzed these data using the same pipelines in earlier Alzheimer's
752 disease datasets (Fig. 4). The significances of the GE changes between treatments were
753 evaluated in a linear regression model using the p-values of the coefficients deviated
754 from zero.

755 To further identify the regulation of the estrogen and androgen using ChIP-seq
756 dataset (Fig. 5e, 6h), we evaluated the binding scores of ESR1 and AR to the promoter
757 regions of the target genes (i.e., +/- 1kb of the transcription start sites, TSS). The binding
758 score of each target gene (i.e., -10*log10[MACS Q-value] from ChIP-seq analysis) was
759 calculated by the ChIP-Atlas [69], an integrative database of public ChIP-seq data. The
760 ratios in Fig. 5e were calculated by the binding scores of each SFs divided by the
761 median binding score of age-associated SFs common in both sexes.

762

763 **Time series and breakpoint analysis**

764 Autoregressive Integrated Moving Average (ARIMA) model is one of the linear
765 models for forecasting chronological trends. For non-stationary data, *auto.arima*()
766 function from R package *forecast* conducts a stepwise procedure to search for the best
767 model using the smallest Akaike Information Criterion (AIC). This procedure is a
768 variation of the Hyndman-Khandakar algorithm [41].

769 We first corrected the TPM/PSI values for each gene/AS event by removing the
770 unknown batch effect (i.e., surrogate variables) except for the sex and age effect *via* R
771 package *limma*, and averaged the values of the samples with the same sex and age.
772 Then, we applied ARIMA models to figure out genes with potential chronological
773 trends during aging in each sex. Furthermore, we calculated the rate of change at each

774 age point and estimated the breakpoints during aging. This model was conducted based
775 on the approach from a previously published study [16] (URL:
776 https://github.com/UcarLab/SexDimorphismNatureCommunications). In brief, the
777 input data was first transformed into z-ordered objects (R package *zoo*) and the best
778 model was stepwise optimized by AIC (*auto.arima* function from R package *Forecast*).
779 We filtered out the genes/AS events with the significant trend for further analysis by
780 non-seasonal difference order higher than zero (D>0) and the sum of AR or MA
781 coefficient higher than zero (AR+MA>0).

782 The rate of change at each age i was evaluated by the differences between bins in
783 $(i - w) \sim i$ and $i \sim (i + w)$ in the following steps. w indicated a series of window
784 spans from 5 to 15 years old. The p-values were first tested by MANOVA of three PCs
785 after reducing dimensionality in the PCA process and then smoothed by LOESS
786 regression with the smoothing bandwidth in the range from 0.25 to 0.75 (with interval
787 0.05). The p-values of each window span were integrated by average -log10
788 transformed p-values and smoothed by LOESS regressions with span = 0.5. The final
789 rate of changes at each age point was defined by the averaged values through all
790 window spans calculated in the previous step. Moreover, we determined the breakpoints
791 as the age points at the global maximums, as well as the local maximums whose
792 distance to the nearest minimum relative to the global maximum was more than 10%.

793

794 **Identification of aging-disturbance genes (ADGs)**

795 We performed the following steps to identify the ADGs that serve as the key
796 contributors to molecular aging (Fig. 6d). We randomly trimmed 20% of chronological
797 genes 200 times and repeated breakpoint analysis in each iteration. Wilcoxon signed-
798 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 ADGs 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 ADGs identification process. Functional enrichment of sex-biased ADGs was based on MsigDB, and the network between GO terms and genes was constructed by Cytoscape [68]. GE patterns during aging were shown by the R package *pheatmap*.

Our results observed clusters of ADGs 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 ADGs 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 ADGs increased at young ages were defined by the mean of fitted expression levels at 31-40 larger than at 20-30.

DATA 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). 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).

REFERENCE

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 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).

18 Consortium, G. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* **369**, 1318-1330 (2020).

19 Oliva, M. *et al.* The impact of sex on gene expression across human tissues. *Science* **369** (2020).

20 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).

21 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).

22 Gallego-Paez, L. M. *et al.* Alternative splicing: the pledge, the turn, and the prestige. *Human genetics* **136**, 1015-1042 (2017).

23 Wang, Z. & Burge, C. B. Splicing regulation: from a parts list of regulatory elements to an integrated splicing code. *Rna* **14**, 802-813 (2008).

24 Baralle, F. E. & Giudice, J. Alternative splicing as a regulator of development and tissue identity. *Nature reviews Molecular cell biology* **18**, 437-451 (2017).

25 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).

26 Karlebach, G. *et al.* The impact of biological sex on alternative splicing. *BioRxiv*, 490904 (2020).

27 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).

28 Bhadra, M., Howell, P., Dutta, S., Heintz, C. & Mair, W. B. Alternative splicing in aging and longevity. *Human genetics* **139**, 357-369 (2020).

29 Lopes-Ramos, C. M. *et al.* Sex differences in gene expression and regulatory networks across 29 human tissues. *Cell reports* **31**, 107795 (2020).

30 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).

31 Ober, C., Loisel, D. A. & Gilad, Y. Sex-specific genetic architecture of human disease. *Nature Reviews Genetics* **9**, 911-922 (2008).

32 Melé, M. *et al.* The human transcriptome across tissues and individuals. *Science* **348**, 660-

665, doi:10.1126/science.aaa0355 (2015).

33 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).

34 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).

35 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).

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

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

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

39 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, doi:10.1093/gerona/gls068 (2012).

40 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).

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

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

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

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

45 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).

46 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).

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

48 Xiong, X. *et al.* Genetic drivers of m6A methylation in human brain, lung, heart and muscle. *Nature Genetics* **53**, 1156-1165 (2021).

49 Jylhävä, J., Pedersen, N. L. & Hägg, S. Biological age predictors. *EBioMedicine* **21**, 29-36 (2017).

50 Horvath, S. & Raj, K. DNA methylation-based biomarkers and the epigenetic clock theory of ageing. *Nature Reviews Genetics* **19**, 371-384 (2018).

51 Meyer, D. H. & Schumacher, B. BiT age: A transcriptome-based aging clock near the theoretical limit of accuracy. *Aging Cell* (2021).

52 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).
53 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).
54 Hannum, G. *et al.* Genome-wide methylation profiles reveal quantitative views of human aging rates. *Molecular cell* **49**, 359-367 (2013).
55 Li, J. *et al.* in *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM).* 676-679 (IEEE).
56 Song, M. *et al.* A review of integrative imputation for multi-omics datasets. *Frontiers in genetics* **11**, 570255 (2020).
57 Zhang, Y. *et al.* Regional variation of splicing QTLs in human brain. *The American Journal of Human Genetics* **107**, 196-210 (2020).
58 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).
59 Dolgalev, I. msigdbr: MSigDB gene sets for multiple organisms in a tidy data format. *R package version* **7** (2020).
60 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).
61 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).
62 Andrews, S. (Babraham Bioinformatics, Babraham Institute, Cambridge, United Kingdom, 2010).
63 Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nature methods* **9**, 357-359 (2012).
64 Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21 (2013).
65 Kuhn, M. & Johnson, K. *Applied predictive modeling*. Vol. 26 (Springer, 2013).
66 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).
67 Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842 (2010).
68 Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome research* **13**, 2498-2504 (2003).
69 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).

ACKNOWLEDGEMENT

The authors want to thank 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. 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. X.L. provided key suggestions for study design and statistical analysis. D.D. and X.L. downloaded the GTEx RNA-seq and phenotype data.

COMPETING INTERESTS

The authors declare that they have no competing interests.

FIGURES

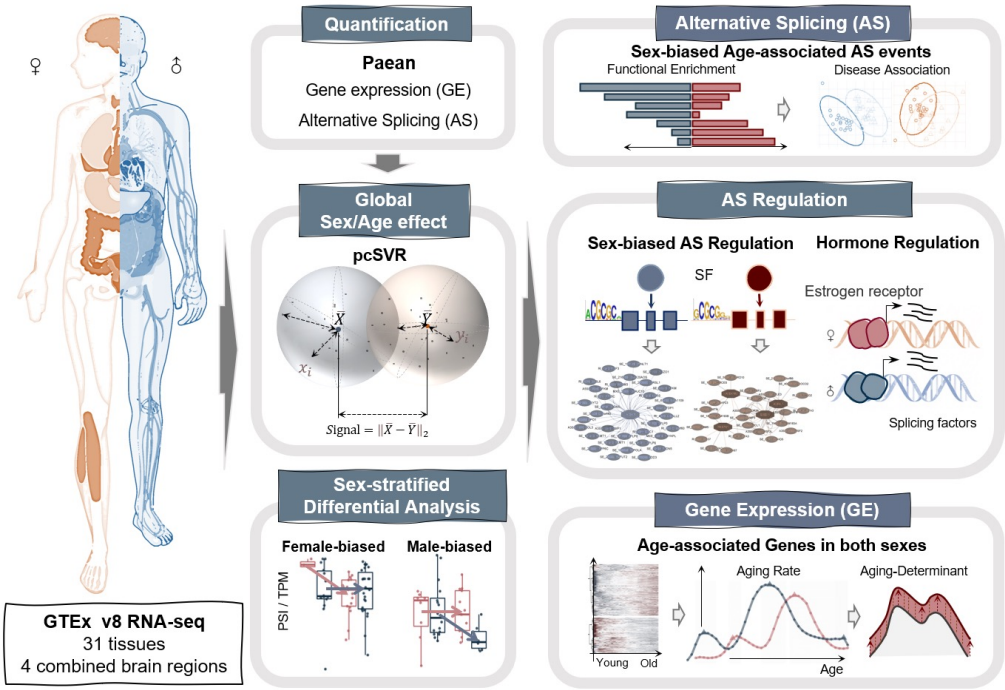

Fig. 1 Schematic Overview of our study.

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 disturbing genes in different sexes.

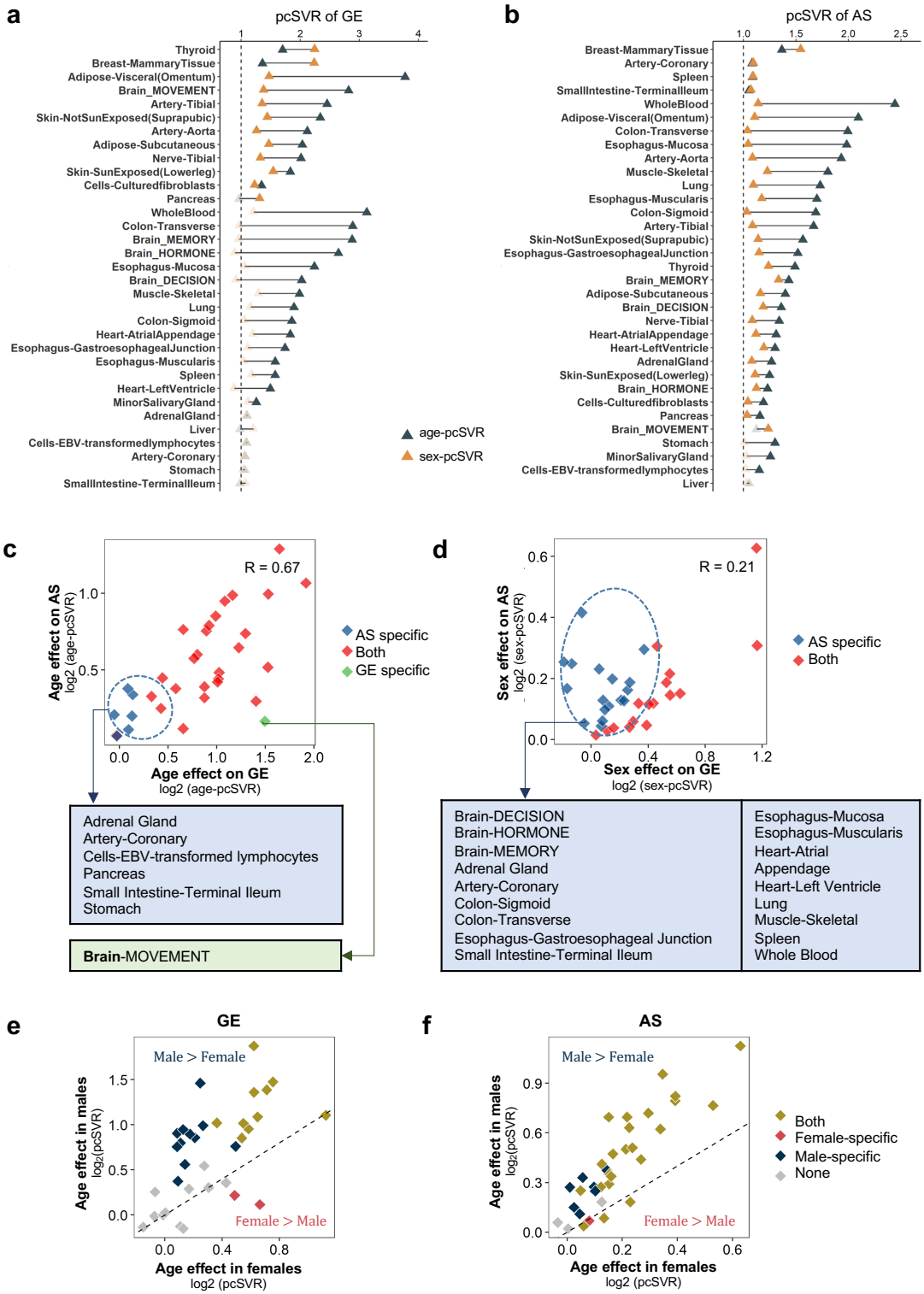

Fig. 2 Individual sex or age effects and combined age-by-sex effects on global transcriptome variation.

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 € 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.

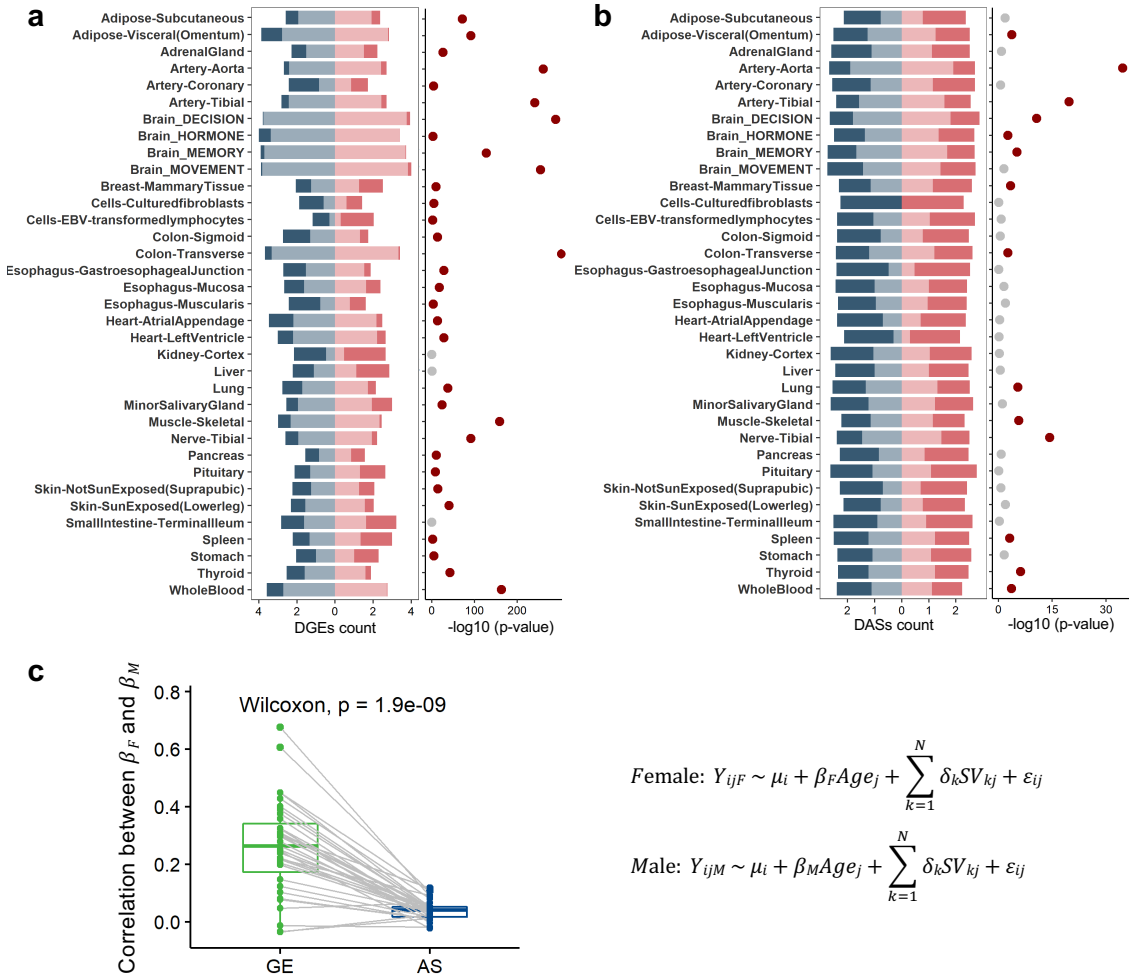

Fig. 3 Sex-stratified age-associated genes and AS events across tissues.

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.

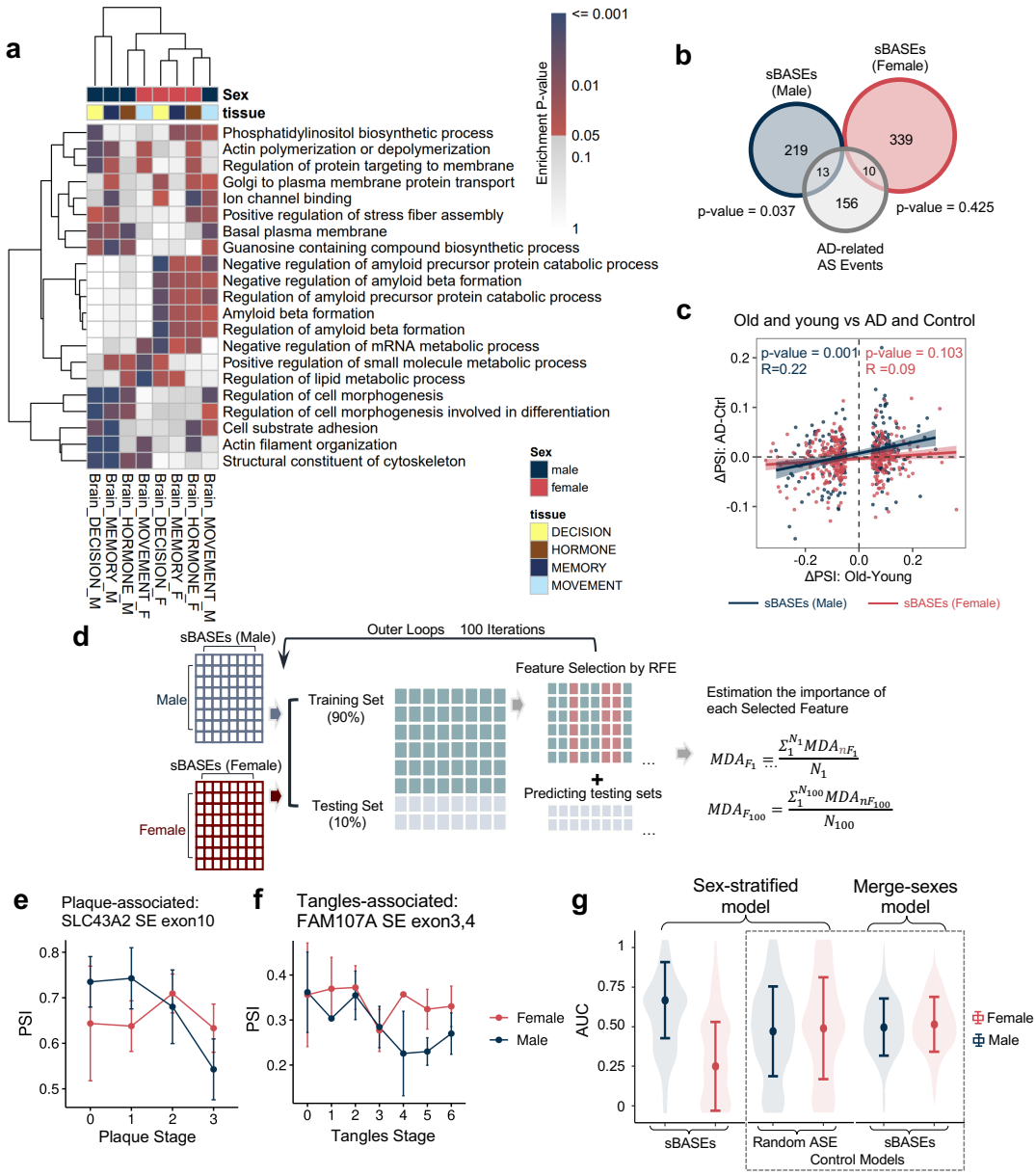

Fig. 4 Male-biased associations between the AS changes during aging and neurodegenerative disease.

a GO analysis of the sBASEs in each sex and brain region. Heatmap shows the sex-specific pathways that are significantly enriched by sBASEs in more than 3 brain regions. Clustering is conducted by default parameters in *pheatmap* functions. Annotations of sexes and brain regions are labeled on the top, and -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, and the other 10% are test sets. This process is repeated 100 times. The recursive feature elimination approach is used for feature selection in each iteration. 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 and 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 ± sd (standard deviation) of AUC in 100 iterations is shown in the boxplot.

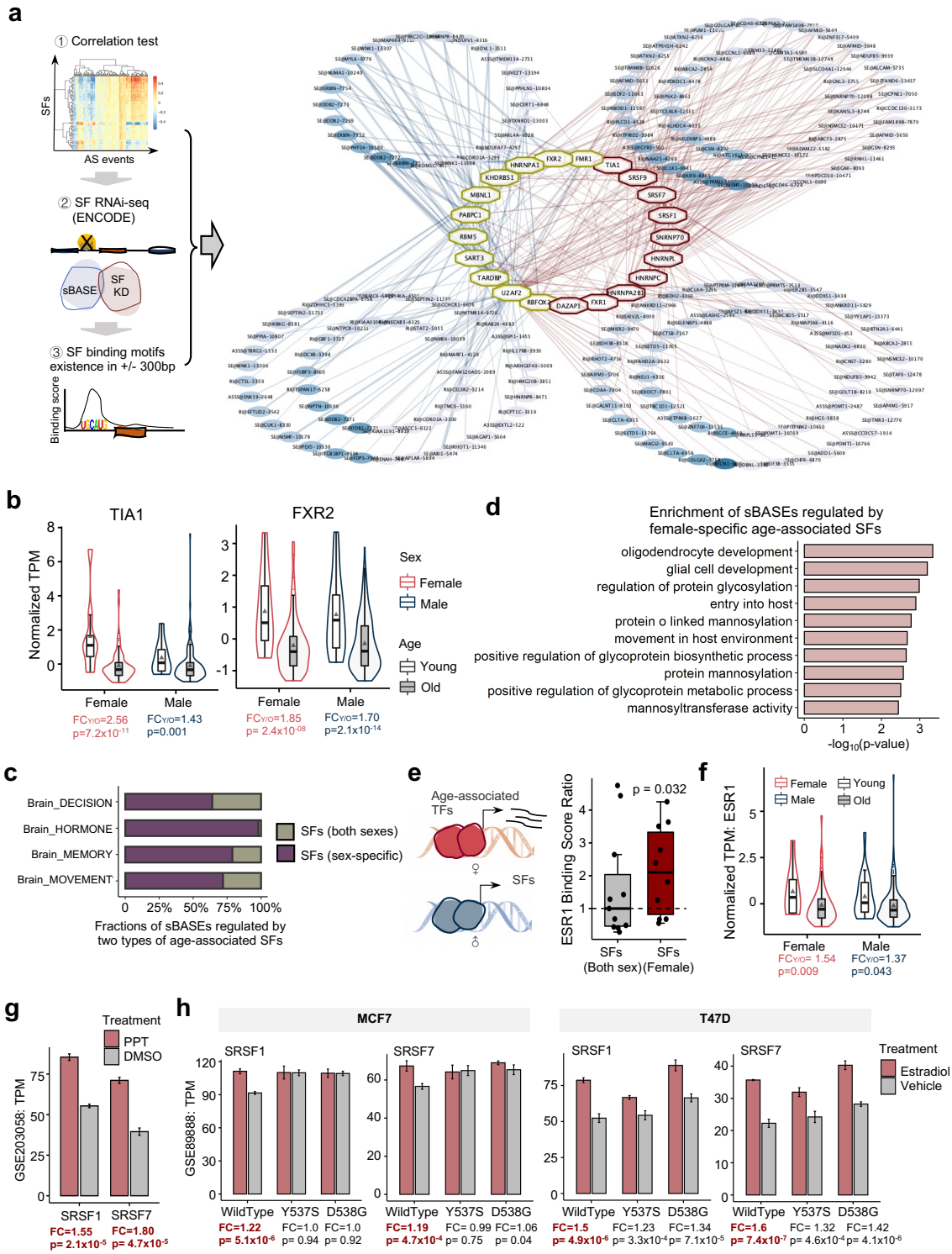

Fig. 5 Sex-dimorphic AS regulation during aging in decision-related brain region. **a** Schematic diagram for constructing AS regulatory networks (left). The networks of the decision-related brain region in females and males are shown on the right. The blue and red lines indicate the regulations in males and females, respectively. Hub genes are SFs and labeled by octagons. The red octagons indicate the female-specific age-associated SFs, while the yellow octagons indicate the age-associated SFs common in

both sexes. The color of the ellipse indicates the significance of sBASEs during aging, and the thickness of the line shows the weight of the edge (-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 Y-axis indicates the normalized TPM values. The p-value and foldchange between old and young groups in each sex are labeled at the bottom.

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. The Y-axis indicates -log10 (enrichment p-value).

e Schematic diagram for the transcriptional regulation of sex hormones on age-associated SFs via nuclear receptors (left). The ESR1 binding scores on the regulatory age-associated SFs are shown on the right. The Y-axis indicates the ratio of ESR1 binding scores of each SFs divided by the median binding score of age-associated SFs common in both sexes.

f Expression levels of ESR1 during aging. The Y-axis indicates the normalized TPM values. The p-value and foldchange between old and young groups in each sex are labeled at the bottom. The legends of each group are the same as those in **b**.

g Expression levels of SRSF1 and SRSF7 treated by estrogen receptor agonist (PPT) *vs.* DMSO in MCF-7 cell line. The Y-axis indicates the TPM values; the p-value and foldchange between PPT and DMSO treatment are labeled at the bottom.

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. Y-axis indicates the TPM values; the p-value and foldchange between estradiol treatment and control are labeled at the bottom.

i Change of expression levels of SRSF1 and SRSF7 during aging, as plotted by mean ± s.e. (standard error).

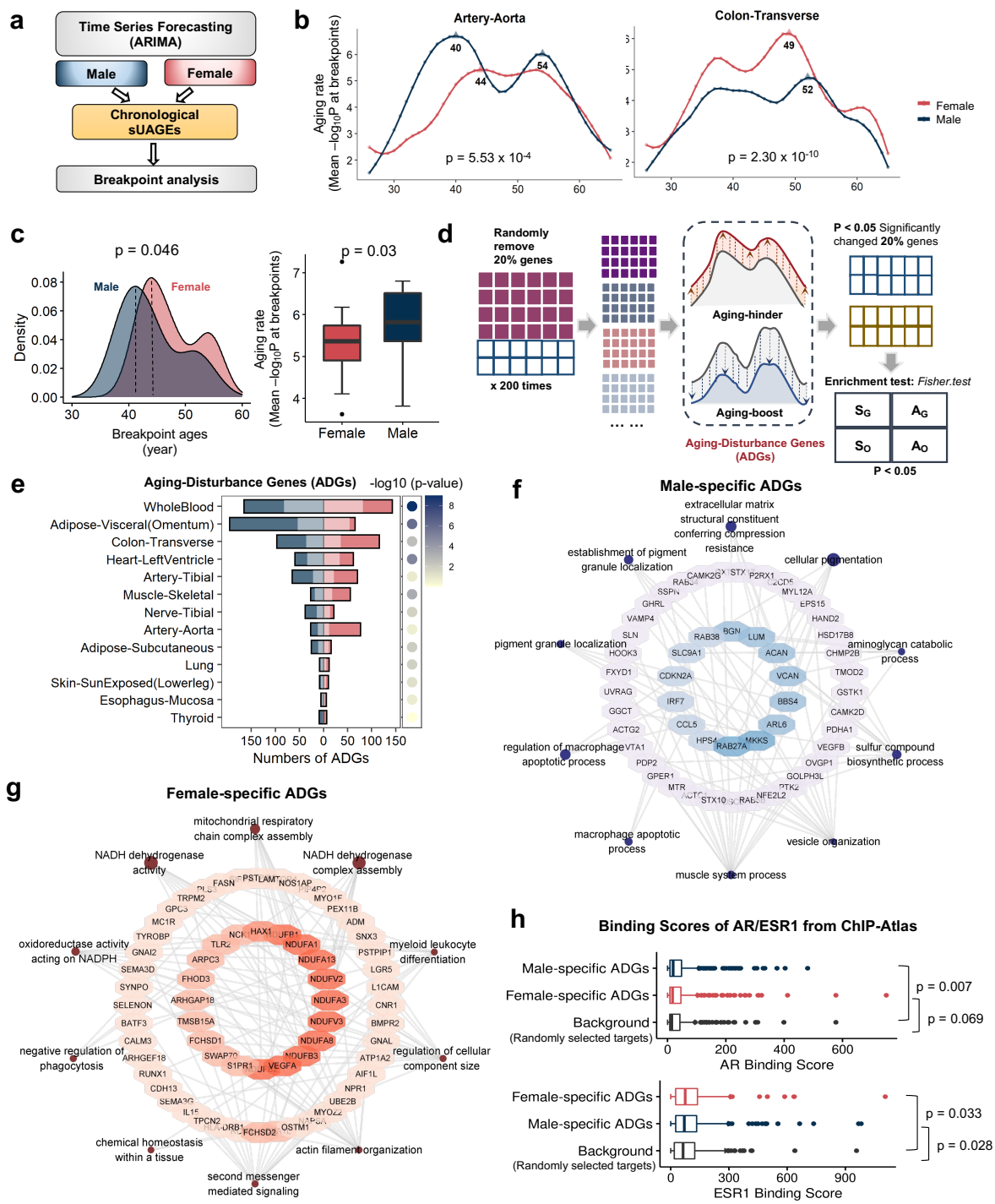

Fig. 6 Sex-dimorphic aging rate of GE during aging process.

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 Methods). The p-values are calculated using Wilcoxon signed-rank test.

c Characteristics of the major breakpoints across tissues in each sex. The left panel shows the distribution of the major breakpoints across multiple tissues. The right panel

shows the aging rate at the major breakpoints. The p-values are calculated using Wilcoxon signed-rank test.

d Pipelines for identifying ADGs in each sex and tissue (see Method).

e ADGs in different sexes and tissues. The numbers of ADGs in females (red) and males (blue) are shown in the barplot (left). Tissues are ordered by the number of ADGs common in both sexes, which are shown in lighter colors. Significances of the overlapped ADGs are calculated using hypergeometric test and colored in a gradient palette (right).

f, g GO analysis of male-biased ADGs (**f**) and female-biased ADGs (**g**). Sex-specific GO terms with top10 -log10 transformed enrichment p-values are plotted. GO terms are labeled by purple dots, the size of which indicates the significance of the enrichment analysis. The larger the dots, the more significant enrichment scores of the biological pathways. ADGs are labeled by blue dots, the color of which indicates the number of GO terms associated with those genes.

h The binding scores of AR (top) and ESR1 (bottom) on male-specific ADGs, female-specific ADGs, and randomly selected AR/ESR1 target genes. The p-values are calculated by Wilcoxon rank-sum test.