

RESEARCH

Open Access



Bioinformatics analysis of diagnostic biomarkers for Alzheimer's disease in peripheral blood based on sex differences and support vector machine algorithm

Wencan Ji^{1,2}, Ke An^{2,3}, Canjun Wang⁴ and Shaohua Wang^{1,2,3*}

Abstract

Background: The prevalence of Alzheimer's disease (AD) varies based on gender. Due to the lack of early stage biomarkers, most of them are diagnosed at the terminal stage. This study aimed to explore sex-specific signaling pathways and identify diagnostic biomarkers of AD.

Methods: Microarray dataset for blood was obtained from the Gene Expression Omnibus (GEO) database of GSE63060 to conduct differentially expressed genes (DEGs) analysis by R software limma. Gene Ontology (GO) analysis, Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis and Gene set enrichment analysis (GSEA) were conducted. Immune checkpoint gene expression was compared between females and males. Using CytoHubba, we identified hub genes in a protein–protein interaction network (PPI). Then, we evaluated their distinct effectiveness using unsupervised hierarchical clustering. Support vector machine (SVM) and ten-fold cross-validation were used to further verify these biomarkers. Lastly, we confirmed our findings by using another independent dataset.

Results: A total of 37 female-specific DEGs and 27 male-specific DEGs were identified from GSE63060 datasets. Analyses of enrichment showed that female-specific DEGs primarily focused on energy metabolism, while male-specific DEGs mostly involved in immune regulation. Three immune-checkpoint-relevant genes dysregulated in males. In females, however, these eight genes were not differentially expressed. SNRPG, RPS27A, COX7A2, ATP5PO, LSM3, COX7C, PFDN5, HINT1, PSMA6, RPS3A and RPL31 were regarded as hub genes for females, while SNRPG, RPL31, COX7C, RPS27A, RPL35A, RPS3A, RPS20 and PFDN5 were regarded as hub genes for males. Thirteen hub genes mentioned above was significantly lower in both AD and mild cognitive impairment (MCI). The diagnostic model of 15-marker panel (13 hub genes with sex and age) was developed. Both the training dataset and the independent validation dataset have area under the curve (AUC) with a high value (0.919, 95%CI 0.901–0.929 and 0.803, 95%CI 0.789–0.826). Based on GSEA for hub genes, they were associated with some aspects of AD pathogenesis.

Conclusion: DEGs in males and females contribute differently to AD pathogenesis. Algorithms combining blood-based biomarkers may improve AD diagnostic accuracy, but large validation studies are needed.

Keywords: Alzheimer's disease, Sex differences, Biomarker, SVM, Cluster analysis, Hub gene

Introduction

Alzheimer's disease (AD) is the most prevalent neurodegenerative disease among the elderly, which is characterized by memory impairment, language, visuospatial skills

*Correspondence: gjjwsh@126.com

¹ Nanjing Medical University, Nanjing 211166, Jiangsu, China
Full list of author information is available at the end of the article



and other cognitive domains decline [1]. It is reported that women have higher overall incidence than men [2–9]. Sexes differ in genetic drivers [10], clinical severity [11], and neuropathological manifestations [12, 13]. Focusing on the sex-specific AD genetic drivers could transform the way treatments are developed and administered and lead to more personalized interventions.

The primary pathologies of AD is associated with neurofibrillary tangles (NFTs), amyloid- β (A β) plaque deposition, inflammation, synaptic alterations, and neurovascular amyloidosis [14]. The sexual dimorphism in brain structure, genetic background, inflammation, gliosis, and immune module are considered as important implications for mechanistic investigation of AD [15, 16]. Men often have a greater brain volume than women, and so are less sensitive to pathological changes [17], such as atrophy, ATP synthase, the mitochondrial proteome, a redox protein, and cytochrome oxidase [18, 19]. Additionally, preclinical studies have demonstrated a significantly relationship between estrogen and soluble A β levels in the brains of wild-type mice after ovariectomy [20]. Sex differences also present in immune modulation. Women usually have stronger neuro-inflammation and neuro-immune response than men [21]. Some sex-specific genes are associated with AD pathogenesis such as amyloid and tau. For instance, females have a stronger correlation between APOE and tau than males [22]. In the GWAS study of CSF AD biomarkers, the female-specific roles played by SERPINB1 in amyloidosis, OSTN, and CLDN16 in tau pathology have been observed [23]. Another research has confirmed that a male-specific ubiquitin-specific peptidase 9 is a positive regulator of MAPT, a protein associated with AD [24]. As a multifactorial disease, biomarkers are crucial for accurate diagnosis, as well as aiding in the understanding of disease mechanisms. Fluid biomarkers may differ based on sex, but only a few studies have focused on this. Further research is needed.

Recently, with the development of the high-throughput sequencing, bioinformatics analysis is widely applied to unveil underlying mechanisms such as biomarkers identification or molecular classification of diseases [25, 26]. Our study is to investigate the critical sex-associated differentially expressed genes (DEGs) and identify novel diagnostic biomarkers. We first downloaded microarray datasets for peripheral blood in AD and control samples from the GEO database. Stratifying the total sample into two groups distinguished by sex. We identified DEGs among males and females separately. Next, series of enrichment analyses, protein–protein interaction (PPI) analysis, an unsupervised hierarchical clustering analysis and support vector machine (SVM) were performed. We identified hub genes in males and females separately.

Combined with sex and age, a 15-gene-based diagnosis model was constructed. This study provides more molecular insights into the sex differences in AD, and identifies candidate biomarkers for diagnosis (Fig. 1).

Materials and methods

Microarray data acquisition and processing

We downloaded two expression profiling datasets from Gene Expression Omnibus (GEO, <http://www.ncbi.nlm.nih.gov/geo>). We filtered the datasets: 1) datasets with AD in Human Expression Profiling Using Arrays; 2) blood samples of AD; 3) each dataset contains at least five samples; 4) the research contains information about the technology and platform used. Finally, two microarray datasets GSE63060 [27] and GSE63061 [27] with were obtained (Table 1).

Data normalization and identification of DEGs in males and females

Illumina expression chips from the bead series were used as the source of all data. LUMI package in R is used for data processing. Log₂ processing is performed on raw data by lumiExpresso function. We used GSE63060 dataset to screen DEGs by comparing AD samples to Control samples in males and females separately. The download data format is MINIML. R package Limma (version: 3.40.2) [28] was used to analyze the mRNA differential expression. Adjusted *P*-values were calculated in order to correct for false positives. “Adjusted *P* < 0.05 and FC (Fold Change) > 1.3 or FC (Fold Change) < – 1.3” were defined as the thresholds for the screening mRNAs for differential expression. The adjusted *P* value for multiple testing was calculated using the Benjamini–Hochberg correction. We constructed volcano plots based on fold-change values and adjusted *P*. The heat map is displayed by the R software package pheatmap. Using a Venn diagram, we identified and visualized the intersecting or sex-specific DEGs for males and females.

Enrichment analysis and immune-checkpoint-relevant gene expression

Gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses were performed in R (version 3.6.3) with clusterProfiler (version 3.14.3) and calculated zscore by GO plot package [29] (version 1.0.2). These enrichment results were visualized using chord plots. Significantly enriched functions and pathways was screened of *p*-value < 0.05. Gene set enrichment analysis (GSEA) was used to analyze the distribution trend of the genes of a predefined set. GSEA was performed by package ggplot2 (Version 3.3.3) of R. The reference gene set is c2.cp.v7.2.symbols.gmt in the MsigDB database (<https://www.gsea-msigdb>).

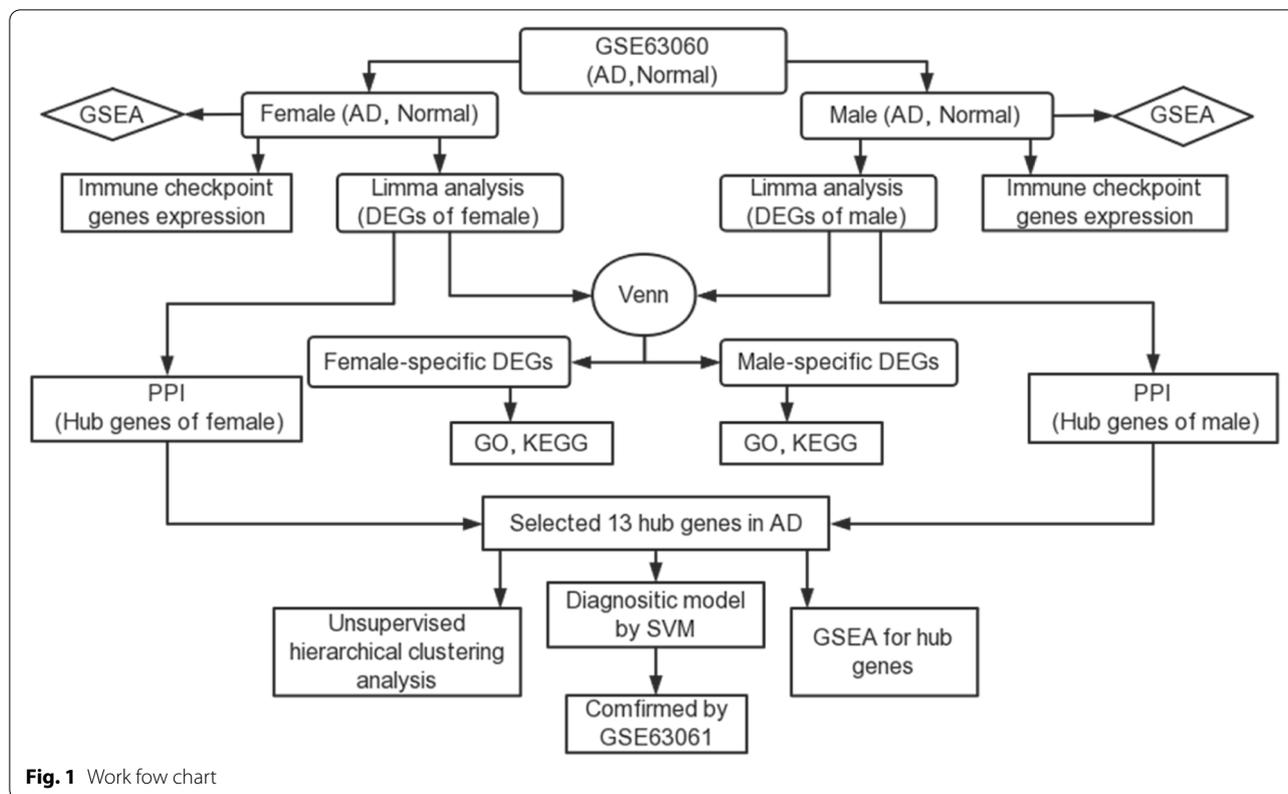


Table 1 Characteristics of the selected microarray datasets

Dataset	Platform	AD: MCI: Control	Age	female: male	Country
GSE63060	GPL6947	145: 78: 102	AD (58-88y) MCI (63-90y) Control (52-87y)	198: 127	United Kingdom
GSE63061	GPL10558	138: 109: 133	AD (59-95y) MCI (57-100y) Control (63-91y)	231: 149	United Kingdom

org/gsea/msigdb/collections.jsp). After 1,000 permutations, genes with an adjusted p -value < 0.05 are statistically significant [30].

Finally, we validated the immune-checkpoint-relevant gene expression in males and females separately. The statistical difference of two groups was compared through the Wilcox test.

PPI network construction and hub gene analysis

To further explore the interactions among DEGs in males and females separately, PPI network analysis was performed using the online tool STRING (<https://string-db.org/>, version 11.5) [31] with a threshold of combination > 0.4. Cytoscape (version 3.8.2) (<https://cytoscape.org/>) was used

to import the interaction information. Cytoscape’s CytoHubba plugin was used to identify the hub genes. The top 20 hub genes were calculated using five algorithms including stress centrality, closeness centrality, radiality centrality, maximum neighborhood component (MNC), and degree. Finally, the hub genes were identified by intersecting the top 20 genes.

Unsupervised hierarchical clustering analysis

To examine the effectiveness of hub genes in distinguishing AD and Control samples, an unsupervised hierarchical clustering analysis was performed. Consensus Cluster Plus R package (version v1.54.0) Clustering was used for clustering, six clusters are the maximum, and a

total of 80% of the samples are drawn 100 times, cluster-*Alg*="hc", inner *Linkage*='ward.D2' [32]. Gene expression heatmaps with *SD*>0.1 are maintained using the R software package *pheatmap* (version 1.0.12).

Classification prediction of AD using SVM

An analysis of classification was conducted using Support Vector Machine (SVM) and Python (version 3.8). To reduce the effects of over fitting, our development and validation of a diagnostic model for AD includes a training step and a validation step. Most importantly, the model was evaluated using an independent data set only once. First, in each task of classification, radial basis function (RBF) were selected as kernel functions and cost (*C*) and gamma (γ) of the kernel function were found to be the best by a grid-search approach using ten-fold cross-validation and establishing receiver operation characteristic (ROC) curves. Finally, this machine learning model was evaluated by classifying AD in a completely independent validation set (GSE63061). Using the area under the receiver operating curve (AUC), the diagnostic performance was estimated. The diagnostic value of SVM classifier model was further verified by building an SVM with confusion matrix.

To recognize the biological process of the 13 hub genes that are possibly associated with AD in GSE63060 datasets, GSEA was performed again.

Results

Identification of DEGs

The dataset GSE63060, which included 99 AD and 60 Control in females and 46 AD and 42 Control in males, were carried out to analyze the DEGs. We identified 113 downregulated DEGs in females (Fig. 2A, B and Supplementary Data 1), and 83 downregulated DEGs along with 20 upregulated DEGs were observed in males (Fig. 2C, D and Supplementary Data 2). Next, volcano plot and heatmap analyses were used to visualize these DEGs.

Identification of sex-specific DEGs in males and females

We intersected DEGs lists of males and females from dataset GSE63060. Finally, 37 female-specific DEGs and 27 (18+9) male-specific DEGs were identified (Fig. 2E and Supplementary Data. 3).

Enrichment analysis

Analyses of GO and KEGG were used to investigate the functionalities of these sex-specific DEGs, respectively. GO analyses provide three different domain of biological processes (BP), molecular function (MF), and cellular component (CC). The top 3 GO analyses were selected and were drawn in a chord plot. Female-specific DEGs were mainly enriched in proton transmembrane

transport, oxidative phosphorylation and ribosome assembly (BP); proton-transporting two-sector ATPase complex, catalytic step 2 spliceosome and mitochondrial respiratory chain(CC); proton transmembrane transporter activity, activity of cysteine-type endopeptidase in apoptosis, activity of proton-transporting ATPase and rotational mechanism (MF) (Fig. 3A and Supplementary Data 4). Male-specific DEGs tended to be enriched in killing of cells of other organism, disruption of cells of other organism and antimicrobial humoral response (BP); cytoplasmic vesicle lumen, vesicle lumen and secretory granule lumen (CC); structural constituent of ribosome, rRNA binding and protease binding (MF) (Fig. 3B and Supplementary Data 5). Next, we showed the KEGG pathways in the column chart. The KEGG pathway enrichment analysis in females showed that the DEGs were enriched in oxidative phosphorylation, protein export and collecting duct acid secretion (Fig. 3C and Supplementary Data 4). The KEGG pathway analysis in males revealed that the DEGs were significantly enriched in ribosome, transcriptional misregulation in cancer, staphylococcus aureus infection and NOD-like receptor signaling pathway (Fig. 3D and Supplementary Data 5).

In order to multi-perspective observation of the enrichment pathway, we will use GSEA to assess the genes' contribution to the phenotype. In both females and males, 21 and 16 pathways were enriched, respectively. Among them, there are 5 female-specific pathway and 2 male-specific pathway (Fig. 3E, F, Supplementary Data 6 and 7).

Immune-checkpoint-relevant gene expression in males and females

Immune checkpoint molecules are regulatory molecules that inhibit the immune system. Eight immune suppressive molecules were selected [33, 34]. Female and male immunocheckpoint-related genes were depicted through boxplots (Fig. 4). As compared with Control samples, HAVCR2 and LAG3 expression levels were significantly higher in male AD samples at 9.978 and 8.373 ($P=0.012$ and 0.041; Wilcoxon rank sum test). CTLA4 was significantly lower in male AD samples at 7.592 ($P=0.032$; Wilcoxon rank sum test). However, these eight genes were not differentially expressed in females (Supplementary Data 10).

PPI network construction and hub gene selection

DEGs in females and males were inputted separately into STRING to obtain PPI records. The interaction network was visualized by Cytoscape 3.8.2 (Fig. 5A, B).

Next, the hub genes were identified with the cytoHubba plugin. According to the five algorithms of Degree, MNC, Radiality, Stress and Closeness, the top 20 hub genes were selected. Finally, eleven hub genes

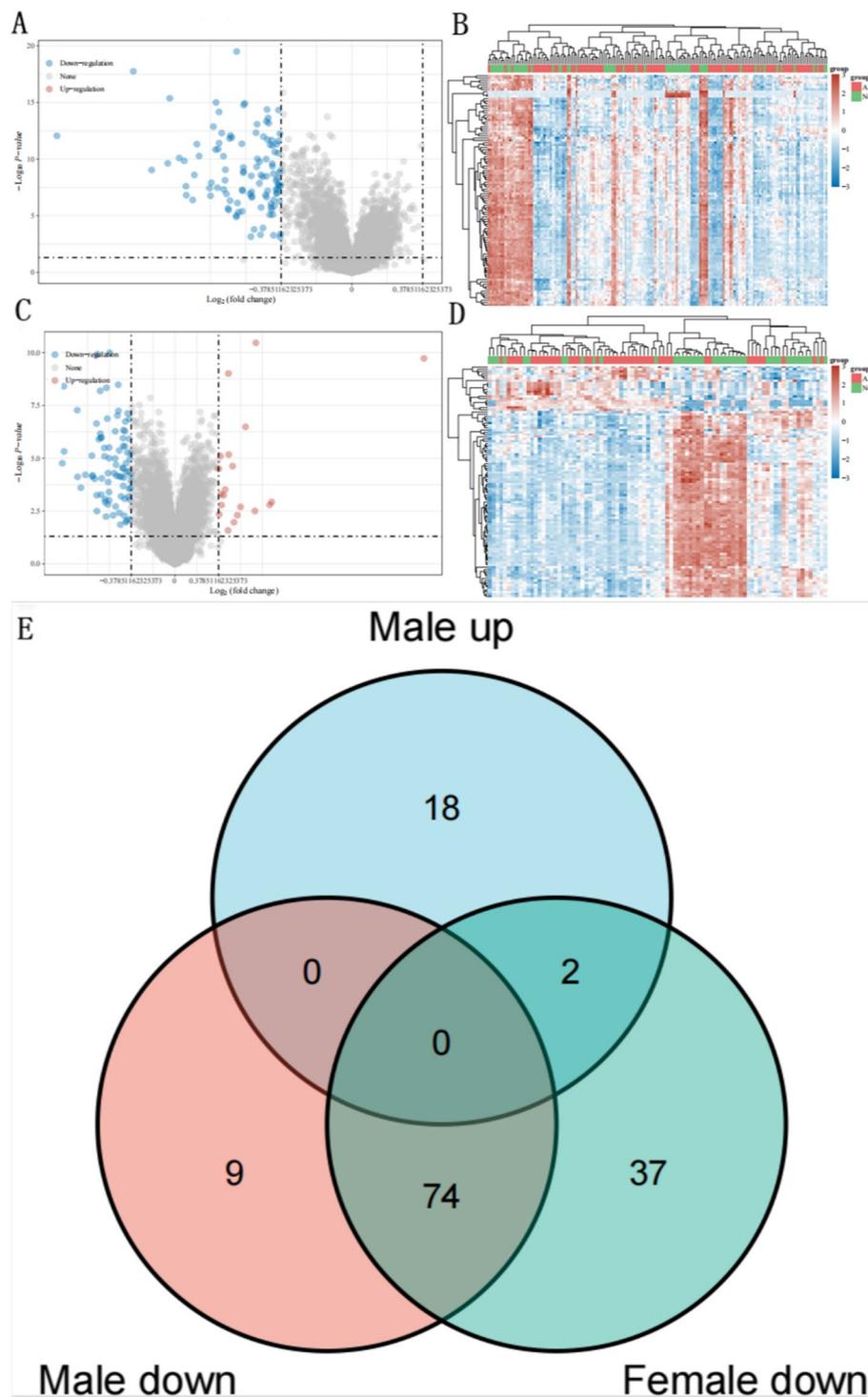


Fig. 2 Identification of DEGs and venn diagram analysis in GSE63060. **A** Volcano plot of females. **B** Heatmap of females. **C** Volcano plot of males. **D** Heatmap of males. Volcano plot: red marks upregulated genes; grey marks non-significant genes; blue marks downregulated genes. Heatmap: red marks high expression; blue marks low expression. **E** Venn diagram of female-specific DEGs and male-specific DEGs. Light green: female-specific DEGs; orange and light blue: male-specific DEGs. AD: Alzheimer’s disease; NC: Normal control

(SNRPG, RPS27A, COX7A2, ATP5PO, LSM3, COX7C, PFDN5, HINT1, PSMA6, RPS3A and RPL31) in females and eight hub genes (SNRPG, RPL31, COX7C, RPS27A, RPL35A, RPS3A, RPS20 and PFDN5) in males were identified (Fig. 5C, D, Supplementary Data 8 and 9).

To further identify the genetic evidence of sex differences, the top 20 genes by degree are shown separately for males and females (Fig. 5E, F). RPL39 is located on the X chromosome in males, while females lack the first 20 genes on the X chromosome.

Hub gene expression in MCI and AD

Considering the remarkable expression changes of these 13 mRNAs (SNRPG, RPS27A, COX7A2, ATP5PO, LSM3, COX7C, PFDN5, HINT1, PSMA6, RPS3A, RPL31, RPL35A and RPS20) in male and female patients with AD separately, they may play a role in AD pathogenesis. Consequently, the 13 promising mRNAs were selected for further study. Using GSE63060, we assessed whether the expression levels of the 13 hub genes differed. Compared with the Control samples, the expression levels of the 13 genes were significantly decreased among the MCI and AD samples ($p < 0.001$; Dunn's test). Additionally, we found that the mRNA expression levels of RPL31, PSMA6 and COX7A2 in the AD samples were higher than those in MCI samples ($p = 0.042$, 0.035 and 0.020; Dunn's test) (Fig. 6 and Supplementary Data 10).

Our next step was to compare the 13 gene expression level between AD and Control samples by enrolling another independent dataset (GSE63061). As indicated in Fig. 7, AD samples show a significant reduction in the mRNA expression of all 13 hub genes ($p < 0.001$, Wilcoxon rank sum test) (Supplementary Data 10).

Since no single mRNA has been shown to have a prominent diagnostic value to date, greater attention has been given to their synergistic effects. To demonstrate the classification ability of the 13 hub genes combination, an unsupervised hierarchical clustering analysis was performed. As shown in Fig. 8, these hub genes demonstrate an excellent capability for distinguishing samples with similar clinical manifestations.

Identification of Potential Biomarkers of AD Using SVM

According to the analyses above, we obtained different hub genes and performed enrichment analysis for

females and males separately. Besides genes, clinical features such as age and sex may also play an important role in pathogenesis. Thus, we used the expression levels of these 13 genes combined with age and sex data as features to investigate their diagnostic value. SVM analysis with a ten-fold cross validation procedure indicated the AUC of the 15-marker panel was 0.919 (95%CI, 0.901–0.929). Furthermore, SVM with confusion matrix analysis also achieved an accuracy of 85.4%, with sensitivity of 91.0% and specificity of 77.5% (Fig. 9A).

A second microarray dataset GSE63061 was collected and analyzed using SVM in order to confirm the classification reliability of the above selected feature genes. As expected, the 15-marker panel can provide an AUC of 0.803 (95%CI, 0.789–0.826) and achieved an accuracy of 69.4%, with sensitivity of 73.2% and specificity of 61.7% (Fig. 9B).

Towards gaining deeper insights into the 13 hub genes, we identified potential BPs between AD and Control using GSEA. They are shown as a mountain map (Fig. 10). Most of these genes have been implicated in cotranslational protein targeting to membrane, establishment of protein localization to endoplasmic reticulum, protein localization to endoplasmic reticulum and translational initiation.

Discussion

AD is a common disorder in the elderly and females are at higher risk of developing AD than males. Previous researches were usually based on CSF or brain tissue that may limit the application. Many studies have demonstrated that the blood can be used to detect disease-related changes [35–37]. Our study recruited blood tissue microarray data, and we integrated the differentially expressed genes in females and males separately.

Base on sex-specific DEGs, the KEGG pathways with smallest p -value enriched in females include oxidative phosphorylation, protein export, and collecting duct acid secretion. KEGG pathways enriched in males include ribosome, transcriptional misregulation in cancer, staphylococcus aureus infection, and NOD-like receptor signaling pathway. We discovered that abnormality of pathway is often associated with energy metabolism in females, whereas in males, the abnormality is primarily related to immune regulation. Females are more

(See figure on next page.)

Fig. 3 GO and KEGG pathway analyses of DEGs and GSEA plot. **A** The chord plot of DEGs in females. BP includes GO:1902600, GO:0006119, GO:0042255; CC includes GO:0016469, GO:0071013, GO:0005746; MF includes GO:0015078, GO:0097153, GO:0046961. **B** The chord plot of DEGs in males. BP includes GO:0031640, GO:0044364, GO:0019730; CC includes GO:0060205, GO:0031983, GO:0034774; MF includes GO:0003735, GO:0019843, GO:0002020. **C** KEGG in females. **D** KEGG in males. **E** GSEA in females. Female-specific pathway includes gap junction (NES = 1.800, p -adj = 0.036), acute myeloid leukemia (NES = 1.693, p -adj = 0.036), chemokine signaling pathway (NES = 1.677, p -adj = 0.040), phosphatidylinositol signaling system (NES = 1.649, p -adj = 0.036), and mapk signaling pathway (NES = 1.608, p -adj = 0.043). **F** GSEA in males. Male-specific pathway includes toll like receptor signaling pathway (NES = 1.875, p -adj = 0.038) and natural killer cell mediated cytotoxicity (NES = 1.819, p -adj = 0.038)

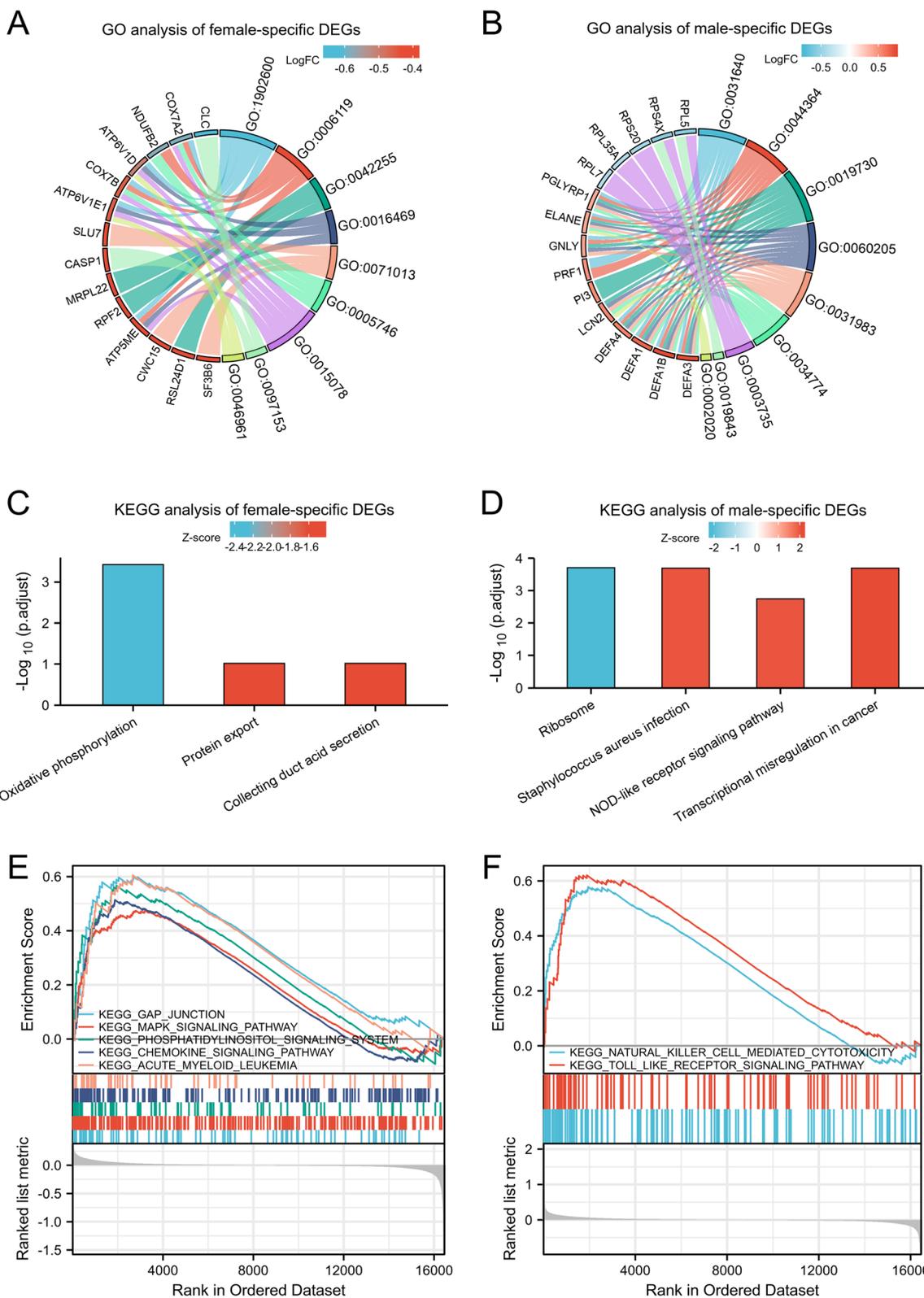
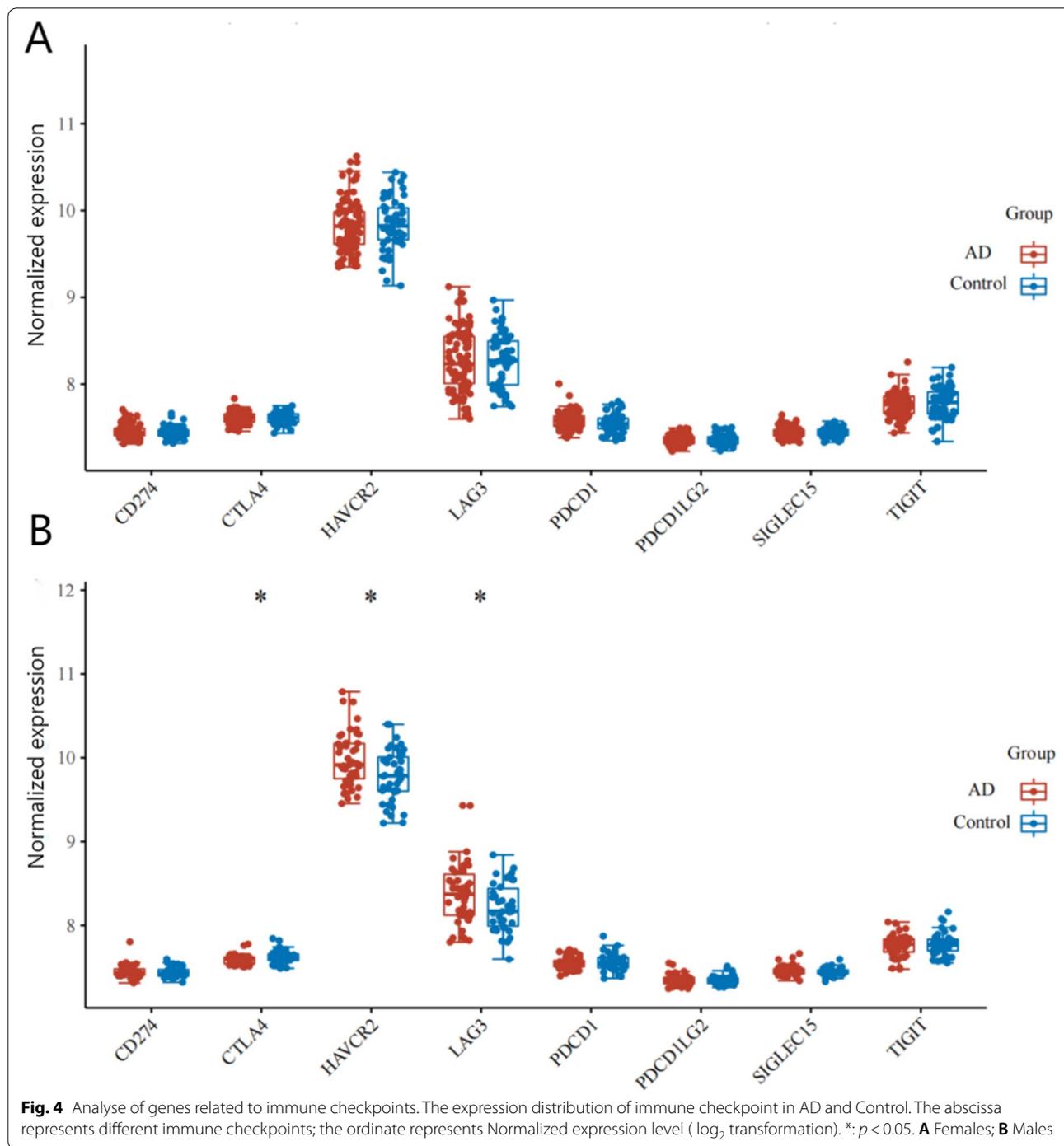


Fig. 3 (See legend on previous page.)



susceptible to estrogen’s effects on energy homeostasis. Research has shown that estrogen in females plays a significant role in AD sex differences in preclinical and clinical studies [38, 39]. Female-specific pathways have also been found in a study of sex differences, but there were no male-specific pathways, possibly because tissue and ethnicities differ [40]. During AD pathogenesis, ribosome

dysfunction can lead to altered translation, specifically in astrocytes [41, 42]. During immune challenge, male astrocytes produced more inflammatory molecules than female derived from cerebral cortex [43]. These studies show that males generally appear to have higher neuroimmune tone in the brain. However, the association between sex and these pathways involved in AD has not

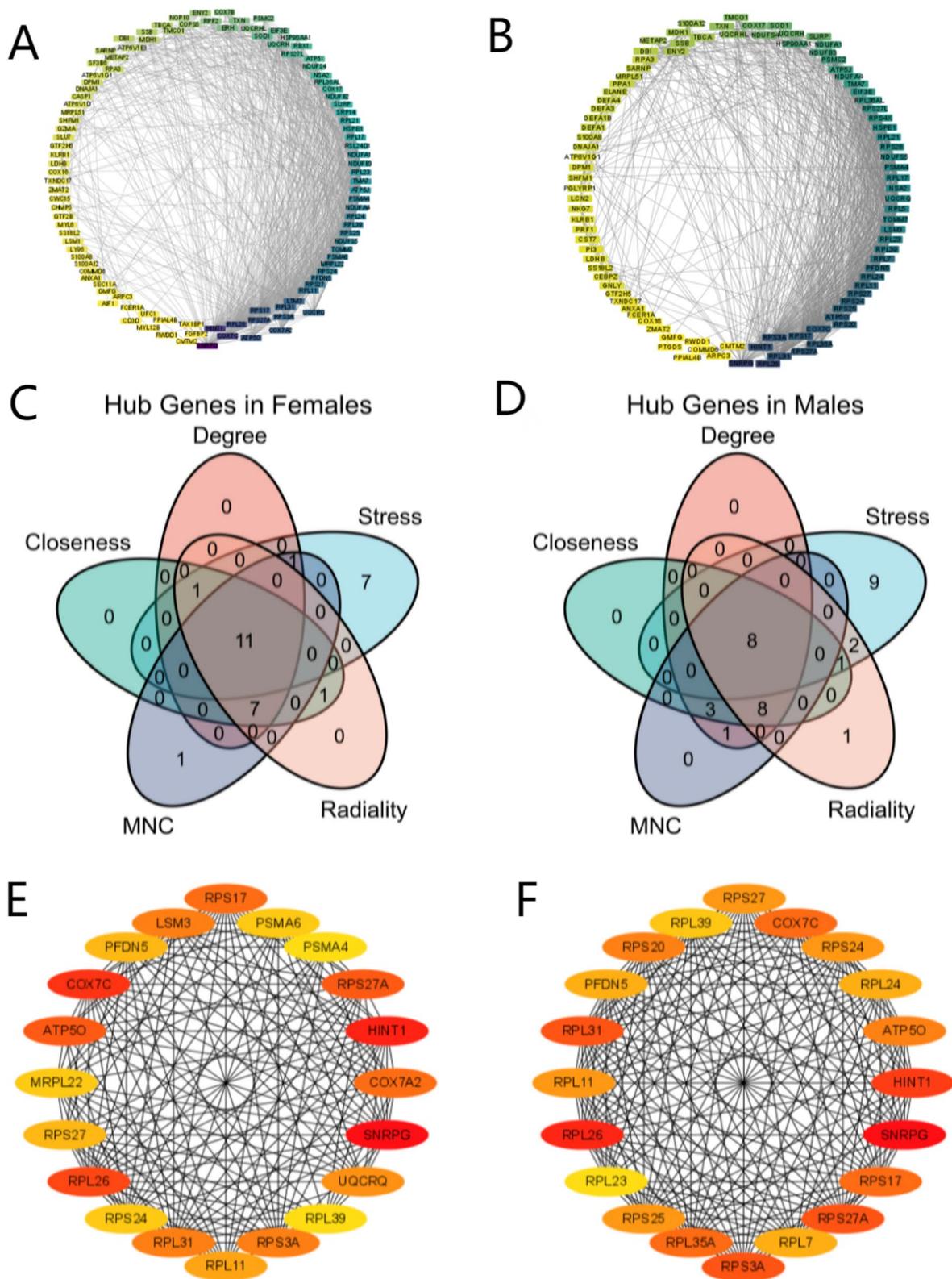
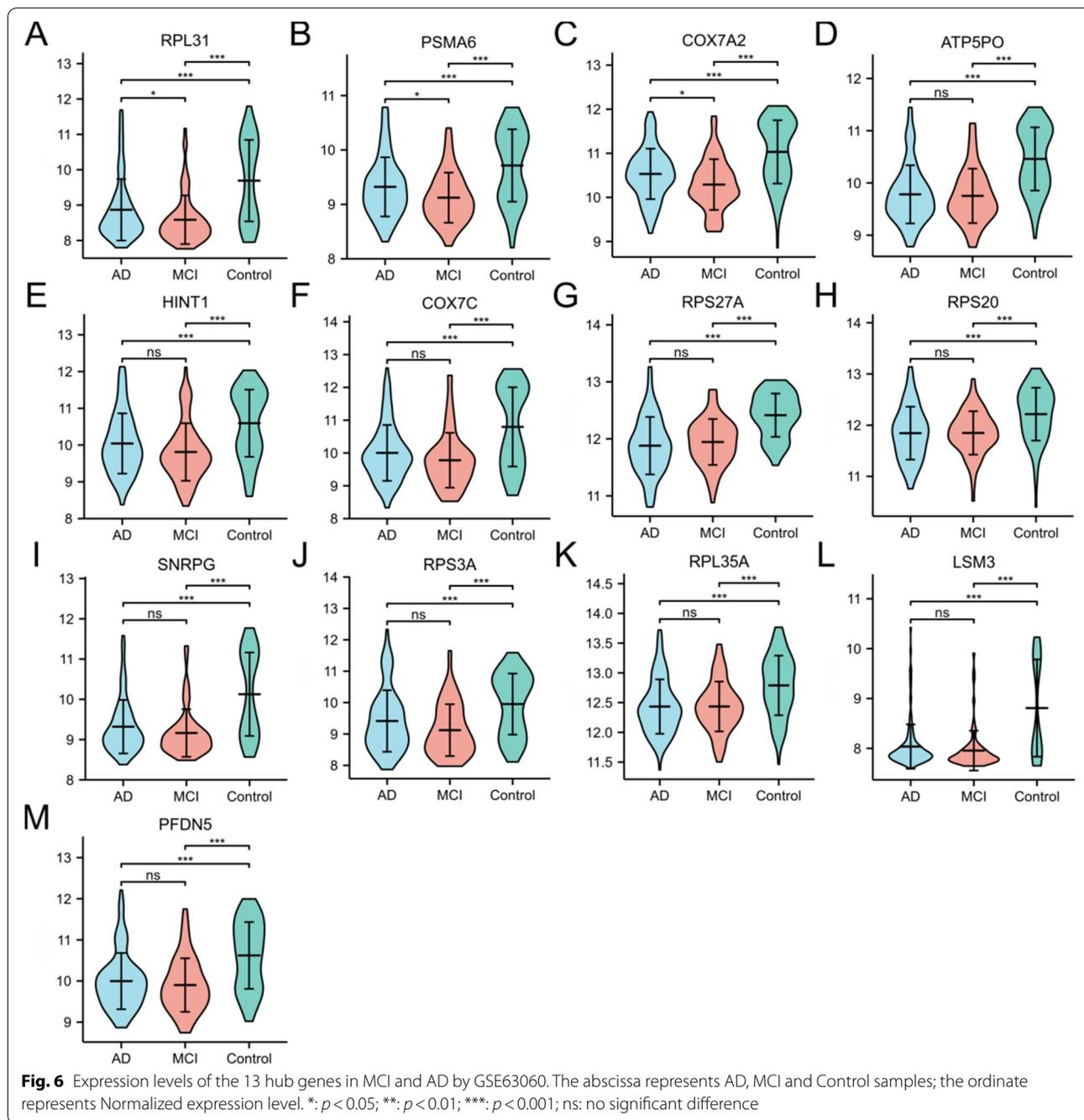


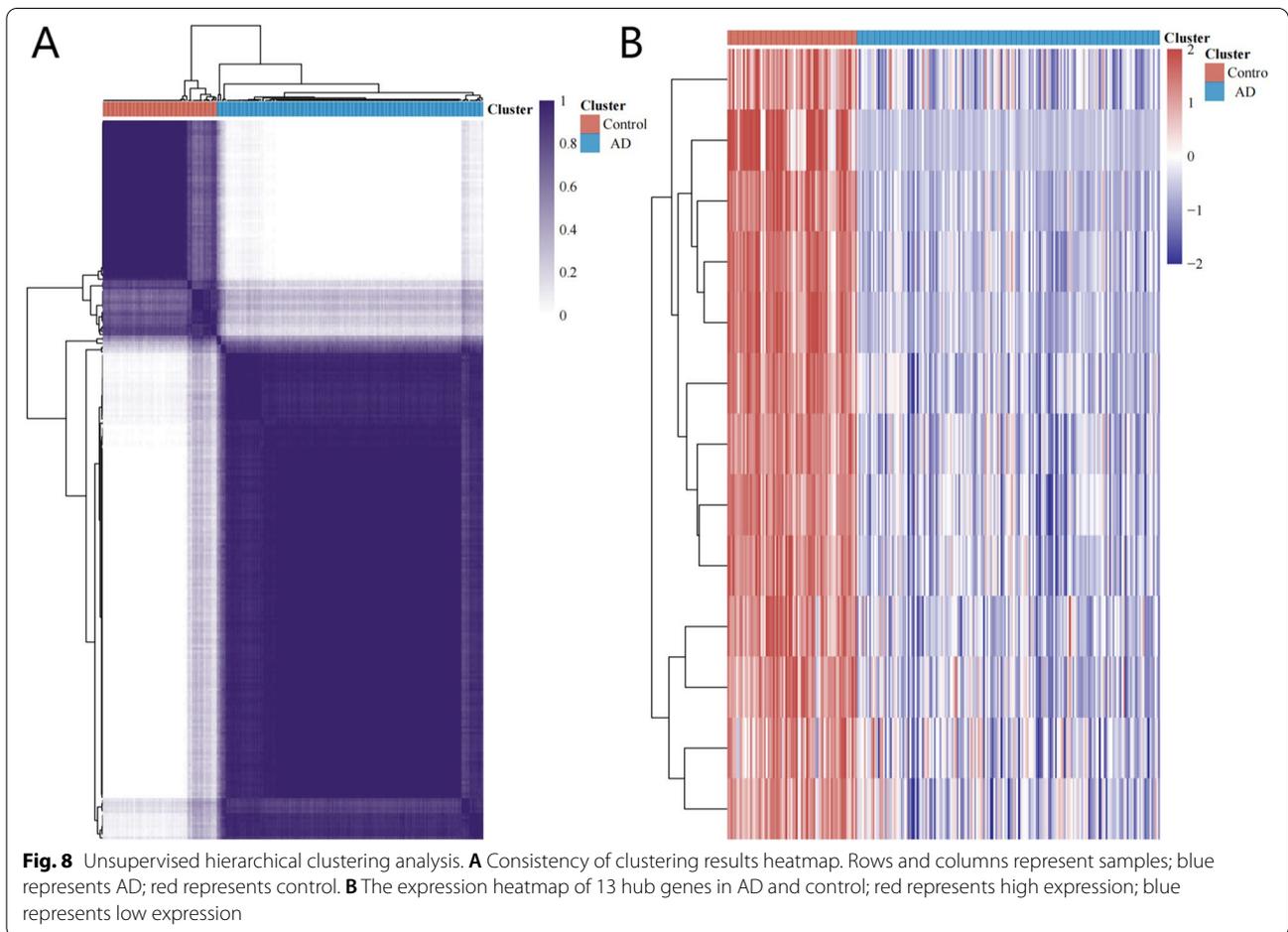
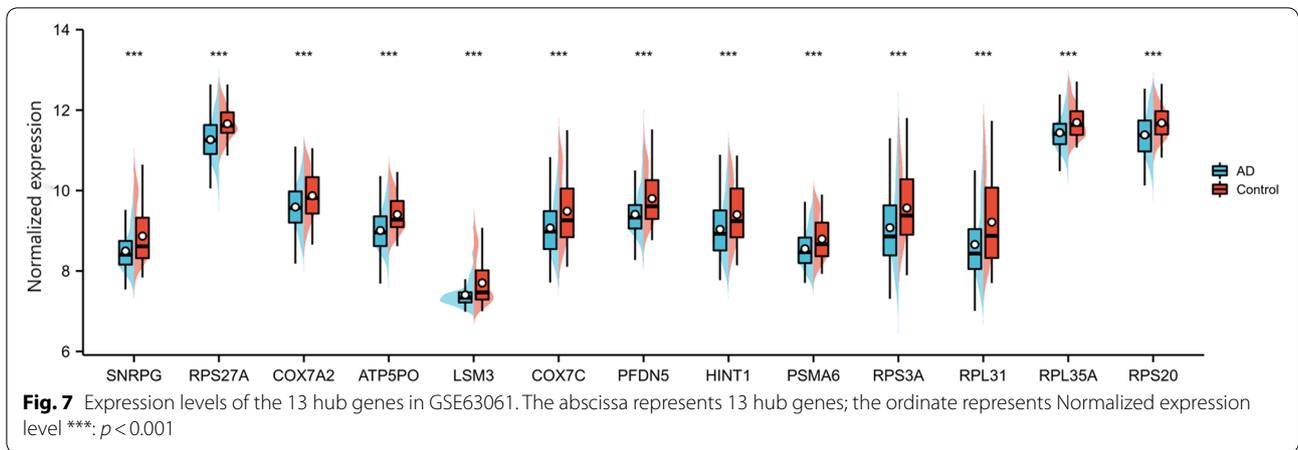
Fig. 5 PPI network construction and hub gene selection. **A** PPI network of DEGs in females. **B** PPI network of DEGs in males. **C** Hub gene selection of females. **D** Hub gene selection of males. **E** The top 20 genes of degree in the female PPI network. **F** The top 20 genes of degree in the male PPI network. MNC: maximum neighborhood component



been studied. Additionally, analyzing the BP of male-specific DEGs showed that immune function, such as killing of cells of other organism, disruption of cells of other organism, antimicrobial humoral response, were stronger in AD samples.

To gain comprehensive insights, GSEA was conducted to identify potential pathways among males and females. As expected, male-specific pathway including toll like receptor signaling pathway and natural killer

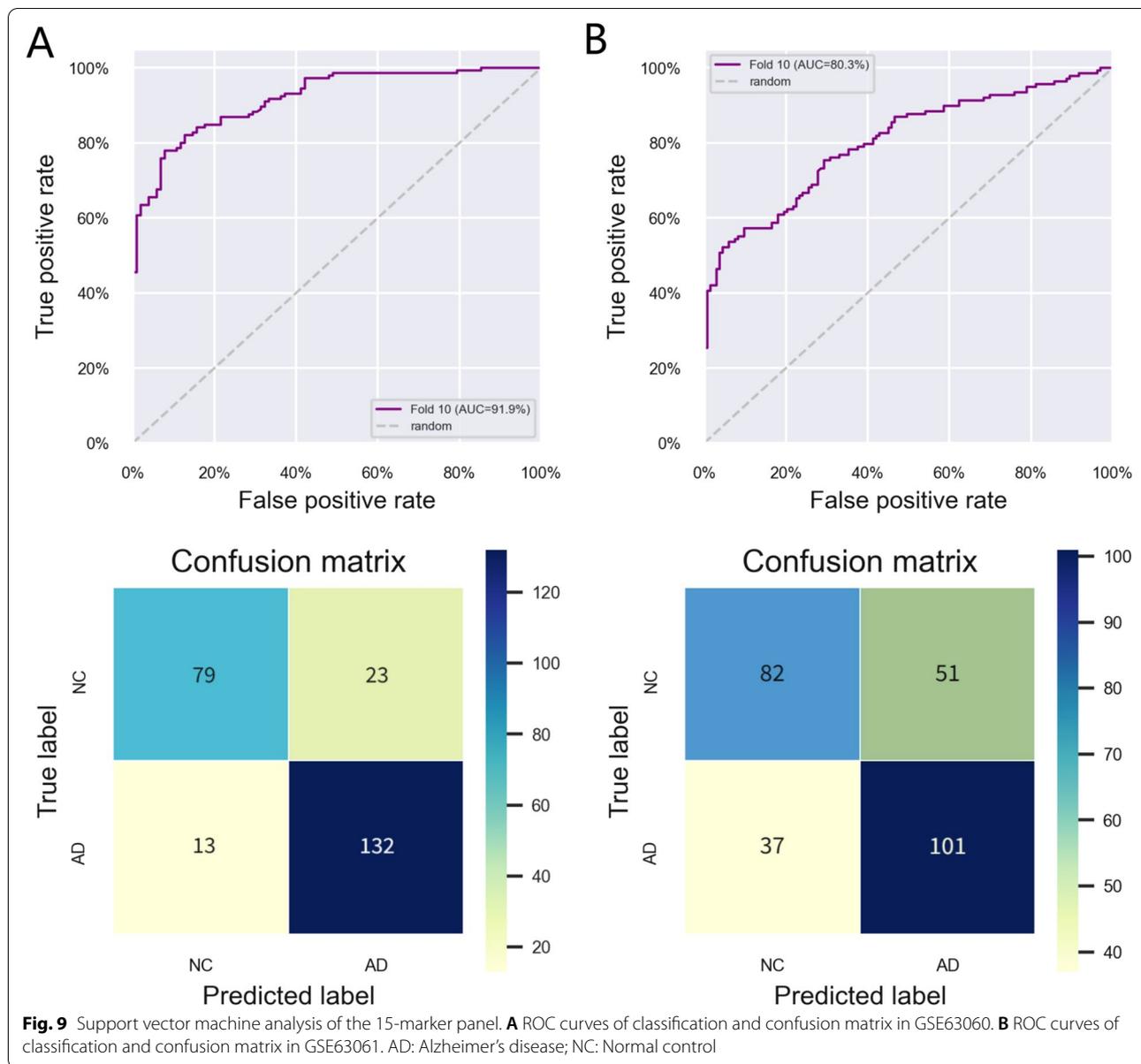
cell mediated cytotoxicity favours an immunological abnormalities. Next, we further detected the immune checkpoint molecules. Three of the eight immune checkpoint genes dysregulated in males, however, there were no differential gene expression in females. An immune checkpoint is a molecule that inhibits the immune system. Through controlling immune responses, they are essential to maintain self-tolerance and prevent auto-immune reactions. The observed sex dimorphism in



immune checkpoint inhibitor responses is likely due to inherent differences between males and females [44–46]. Recent studies confirm that the immune checkpoint contributes to AD pathogenesis. It is well documented that mini-mental state examination (MMSE), tau proteins and Amyloid- β in AD were demonstrated to correlate

with CTLA-4⁺T cells [47]. There could be sex-based differences in AD-intrinsic features that have been shaped by the immune system during AD development and immunoediting.

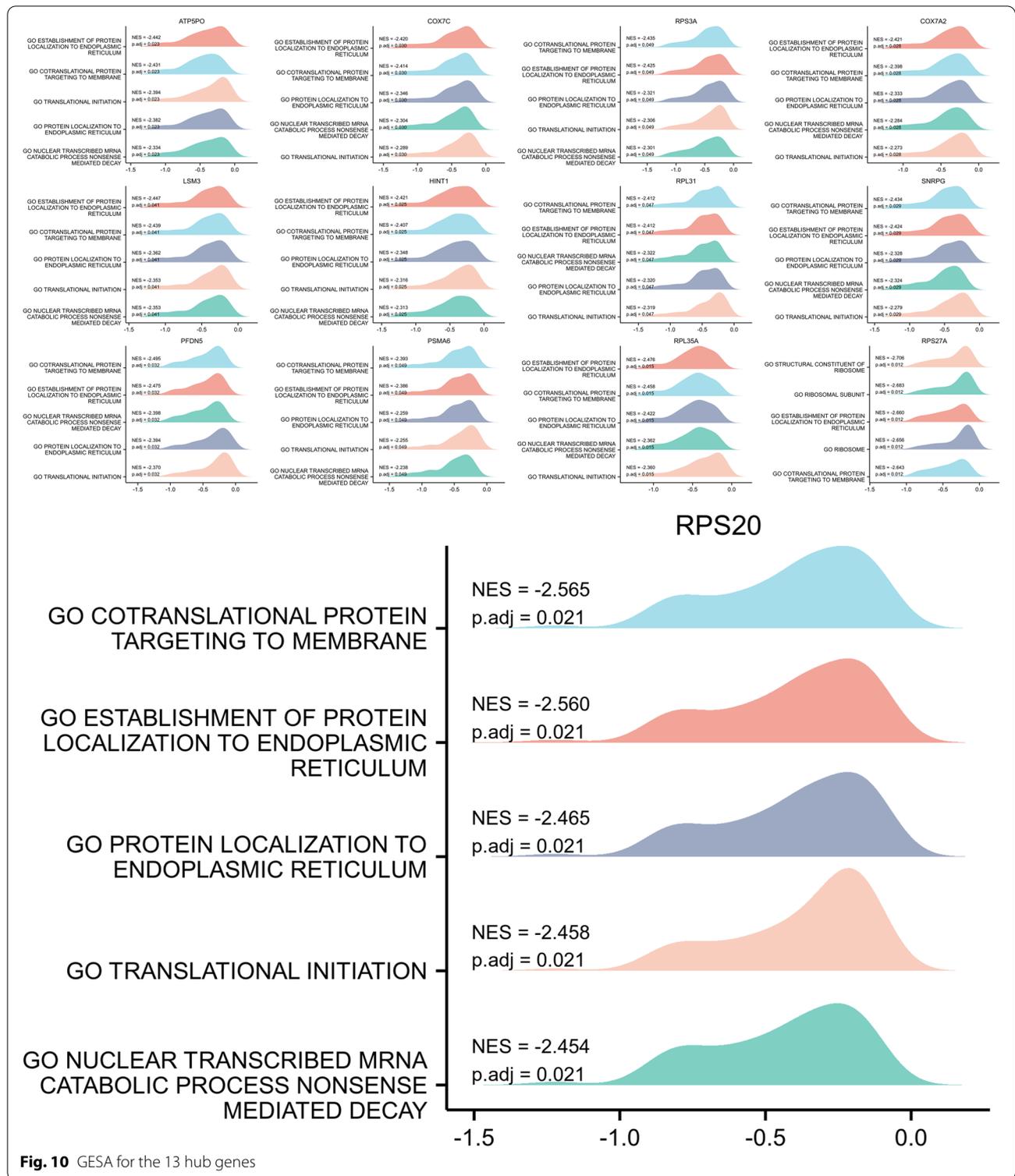
In the PPI network, SNRPG, RPS27A, COX7A2, ATP5PO, LSM3, COX7C, PFDN5, HINT1, PSMA6,



RPS3A and RPL31 were identified as the female-specific hub genes, and SNRPG, RPL31, COX7C, RPS27A, RPL35A, RPS3A, RPS20 and PFDN5 were identified as the male-specific hub genes. SNRPG, RPS27A, COX7C, PFDN5, RPS3A and RPL31 were identified as common hubs. Nevertheless, other hub genes differ greatly between sexes. Moreover, the 13 genes were significantly dysregulated with AD progression.

A cluster of genes associated with the function and structural component of ribosomes, including RPS27A, RPL31, RPL35A, RPS3A, and RPS20, makes up most of the male-specific genes. Transcriptional changes in ribosome-related genes affect downstream translation

in a broader way. Ribosome profiling was utilized in an in vitro experiment to investigate how the lack of oxygen and glucose immediately changes the transcription and translation of the genes in brain cells. The study concluded that oxidative stress has a greater effect on translation than transcription. It is alarming that testosterone can increase oxidative stress [48, 49] and oxidative stress-related conditions [50], and oxidative stress exacerbates testosterone's negative impact on cognition through its negative effects on androgens [51–54]. It can be inferred that sexually dimorphic differences may be a result of the expression of sex-specific ribosomal differential genes in AD. PSMA6 belongs



to component of the 20S core proteasome complex that participates in the ATP-dependent degradation of ubiquitinated proteins. Ubiquitin proteasome system

activity is integral to estradiol's effects on memory, then this could lead to exciting new avenues of basic research into hormonal regulation of cognition, which

could lead to important clinical implications for treating neurodegenerative disorders in which sex-based differences play a role [55]. LSM3 and SNRPG take role in the spliceosome's construction. As a component of the U4/U6-U5 tri-snRNP and the U1, U2, U4, and U5 snRNP complexes, they mostly performs pre-mRNA splicing-related tasks, including as RNA silencing and destruction. A snRNA(U1) accumulation has been observed in Alzheimer's disease [56]. In human brains, identifying cryptic splicing errors with neurofibrillary tangle burden implicates spliceosome disruption and transcriptome perturbation in AD Tau-mediated neurodegeneration [57]. These results show an independent function of U1 snRNA in regulating RNA splicing, suggesting that aberrant RNA processing may mediate neurodegeneration [58]. One study profiled the expression of snRNAs by applying small RNA sequencing to snRNA isolated from anterior cingulate cortex of schizophrenia patients. Two snRNAs were found to be differentially expressed between female cases and controls [59]. However, no report has been published on the relationship between snRNA and sex differences in AD. HINT1 belongs to the histidine triad superfamily, which is widely expressed in different tissues. Ageing of the brain is a major risk factor for many neurodegenerative disorders including Alzheimer's. The downregulation of HINT1 has also been reported in diabetes and AD [60, 61]. Yu W et al. [62] found that potassium 2-(1-hydroxypentyl)-benzoate can increase HINT1 expression levels, thus improving spatial learning and memory deficits in diabetic animals. Chen Q et al.'s study suggests that HINT1 may be associated with schizophrenia and the association is sex specific [63]. According to our study, HINT1 was found to be significantly differentially expressed in female AD. All these results suggest that HINT1 may play a role in neuronal function, but its exact physiological and cellular functions in AD remain unknown. PFDN5 belongs to prefoldin family. It is highly expressed in neurons and other neural cells, which can protect cells from apoptosis by decreases the toxicity of misfolded proteins. Interestingly, whole blood mRNA expression data from Alzheimer's patients revealed downregulation of PFDN5, which could be related to higher levels of toxicity of A β [64]. The results of our research analysis support this conclusion. The present study identified ATP5PO and COX7A2 as female-specific hub genes, which were significantly associated with ATP synthesis, heat production and oxidative phosphorylation. In female brains, estrogen control of glucose metabolism is dismantled during midlife, resulting in a shift in fuel systems and the emergence of dynamic neuroimmune phenotypes.

As fuel use shifts, white matter is at risk of catabolism [65]. PSMA6 belongs to component of the 20S core proteasome complex that participates in the ATP-dependent degradation of ubiquitinated proteins. Ubiquitin proteasome system activity is integral to estradiol's effects on memory, then this could lead to exciting new avenues of basic research into hormonal regulation of cognition, which could lead to important clinical implications for treating neurodegenerative disorders in which sex-based differences play a role [55]. From above analysis, it can be concluded that 13 genes were potential biomarkers for pathogenesis of AD. Assessing and confirming the clinical utility, we constructed a 15-marker panel (13 genes combined with age and sex) based SVM model, which can effectively predicts AD in independent testing sets.

Our work differs most significantly from previous work in the following ways: 1) Here, for the first time, we show the molecular mechanisms of sex differences in peripheral blood from Alzheimer's patients. 2) To identify the diagnostic value of these genes, we used machine learning methods to reduce overfitting and it can be improved in future studies.

Our study has some limitations: 1) Although we registered a second dataset for external verification, our results need to be confirmed by larger-scale studies. 2) The group differences in education and body mass index are potential confounds that could have affected our results, so further research is needed to examine the risks for these subgroups. 3) The mechanisms of these hub genes remain undefined in AD and they could be further explored to elucidate the functions and underlying mechanisms.

Conclusion

Our analyses revealed different hub genes of AD by the analysis of DEGs for males and females separately. We find that the pathophysiological pathways of AD differ in males and females. Using 13 genes as a base, we developed a diagnostic model with a high AUC value in peripheral blood. This study provides insight into the underlying molecular mechanisms for sex dimorphism as well as potential biomarkers that may be useful for diagnostics and therapy.

Abbreviations

AD: Alzheimer's disease; GEO: Gene Expression Omnibus; DEGs: Differentially expressed genes; SVM: Support Vector Machine; GO: Gene Ontology; KEGG: Kyoto Encyclopedia of Genes and Genomes; MCI: Mild cognitive impairment; CSF: Cerebrospinal fluid; BP: Biological Process; CC: Cellular Component; MF: Molecular Function; PPI: Protein-protein interaction; MNC: Maximum Neighborhood Component; GSEA: Gene set enrichment analysis; AUC: Area under curve.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s41065-022-00252-x>.

Additional file 1: Supplementary Data 1. List of DEGs in females.

Additional file 2: Supplementary Data 2. List of DEGs in males.

Additional file 3: Supplementary Data 3. Intersected DEGs lists of males and females.

Additional file 4: Supplementary Data 4. Significantly enriched GO entries and KEGG pathways in females.

Additional file 5: Supplementary Data 5. Significantly enriched GO entries and KEGG pathways in males.

Additional file 6: Supplementary Data 6. GSEA in females.

Additional file 7: Supplementary Data 7. GSEA in males.

Additional file 8: Supplementary Data 8. Degree, MNC, Radiality, Stress and Closeness were used to screen 11 overlapping hub genes in females.

Additional file 9: Supplementary Data 9. Degree, MNC, Radiality, Stress and Closeness were used to screen 8 overlapping hub genes in males.

Additional file 10: Supplementary Data 10. A statistical analysis of immune checkpoint genes and 13 hub genes.

Acknowledgements

Not applicable.

Authors' contributions

Wencan Ji designed the study and extracted the data. Wencan Ji, Ke An and Canjun Wang contributed statistical analyses. Wencan Ji and Shaohua Wang prepared figures and tables. Wencan Ji prepared the manuscript. Shaohua Wang approved the final draft. The author(s) read and approved the final manuscript.

Funding

This work was supported by the National Natural Science Foundation of China (Nos. 81570732 and 81870568, SW).

Availability of data and materials

The datasets generated and/or analyzed during the current study are available in the [GSE63060] repository [<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE63060>], in [GSE63061] repository [<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE63061>].

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

All the authors have consented for the publication.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Nanjing Medical University, Nanjing 211166, Jiangsu, China. ²Department of Endocrinology, Affiliated Zhongda Hospital of Southeast University, Nanjing 210009, Jiangsu, China. ³School of Medicine, Southeast University, Nanjing 210009, Jiangsu, China. ⁴Department of Laboratory Medicine, Affiliated Zhongda Hospital of Southeast University, Nanjing 210009, Jiangsu, China.

Received: 18 May 2022 Accepted: 22 September 2022

Published online: 04 October 2022

References

- Masters CL, Bateman R, Blennow K, Rowe CC, Sperling RA, Cummings JL. Alzheimer's disease. *Nat Rev Dis Primers*. 2015;1:15056.
- Burke SL, Hu T, Fava NM, Li T, Rodriguez MJ, Schuldiner KL, et al. Sex differences in the development of mild cognitive impairment and probable Alzheimer's disease as predicted by hippocampal volume or white matter hyperintensities. *J Women Aging*. 2019;31(2):140–64.
- Dumitrescu L, Barnes LL, Thambisetty M, Beecham G, Kunkle B, Bush WS, et al. Sex differences in the genetic predictors of Alzheimer's pathology. *Brain*. 2019;142(9):2581–9.
- Schumacher Dimech A, Santucci Chadha A, Baracchi F, Girouard H, Misoch S, Giacobini E, et al. Sex differences in Alzheimer disease - the gateway to precision medicine. *Nat Rev Neurol*. 2018;14(8):457–69.
- Gür E, Fertan E, Kosel F, Wong AA, Balci F, Brown RE. Sex differences in the timing behavior performance of 3xTg-AD and wild-type mice in the peak interval procedure. *Behav Brain Res*. 2019;15(360):235–43.
- Koran MEI, Wagener M, Hohman TJ. Sex differences in the association between AD biomarkers and cognitive decline. *Brain Imaging Behav*. 2017;11(1):205–13.
- Mosconi L, Berti V, Quinn C, McHugh P, Petrongolo G, Varsavsky I, et al. Sex differences in Alzheimer risk: Brain imaging of endocrine vs chronologic aging. *Neurology*. 2017;89(13):1382–90.
- Tensil M, Hessler JB, Gutsmiel M, Riedl L, Grimmer T, Diehl-Schmid J. Sex Differences in Neuropsychological Test Performance in Alzheimer's Disease and the Influence of the ApoE Genotype. *Alzheimer Dis Assoc Disord*. 2018;32(2):145–9.
- Toro CA, Zhang L, Cao J, Cai D. Sex differences in Alzheimer's disease: Understanding the molecular impact. *Brain Res*. 2019;15(1719):194–207.
- Neu SC, Pa J, Kukull W, Beekly D, Kuzma A, Gangadharan P, et al. Apolipoprotein E Genotype and Sex Risk Factors for Alzheimer Disease: A Meta-analysis. *JAMA Neurol*. 2017;74(10):1178–89.
- Henderson VW, Buckwalter JG. Cognitive deficits of men and women with Alzheimer's disease. *Neurology*. 1994;44(1):90–6.
- Apostolova LG, Dinov ID, Dutton RA, Hayashi KM, Toga AW, Cummings JL, et al. 3D comparison of hippocampal atrophy in amnesic mild cognitive impairment and Alzheimer's disease. *Brain*. 2006;129(Pt 11):2867–73.
- Hua X, Hibar DP, Lee S, Toga AW, Jack CR Jr, Weiner MW, et al. Sex and age differences in atrophic rates: an ADNI study with n=1368 MRI scans. *Neurobiol Aging*. 2010;31(8):1463–80.
- Guo T, Zhang D, Zeng Y, Huang TY, Xu H, Zhao Y. Molecular and cellular mechanisms underlying the pathogenesis of Alzheimer's disease. *Mol Neurodegener*. 2020;15(1):40.
- Fisher DW, Bennett DA, Dong H. Sexual dimorphism in predisposition to Alzheimer's disease. *Neurobiol Aging*. 2018;70:308–24.
- Zhu D, Montagne A, Zhao Z. Alzheimer's pathogenic mechanisms and underlying sex difference. *Cell Mol Life Sci*. 2021;78(11):4907–20.
- Mielke MM, Vemuri P, Rocca WA. Clinical epidemiology of Alzheimer's disease: assessing sex and gender differences. *Clin Epidemiol*. 2014;6:37–48.
- Ardekani BA, Convit A, Bachman AH. Analysis of the MIRIAD Data Shows Sex Differences in Hippocampal Atrophy Progression. *J Alzheimers Dis*. 2016;50(3):847–57.
- Gallart-Palau X, Lee BS, Adav SS, Qian J, Serra A, Park JE, et al. Gender differences in white matter pathology and mitochondrial dysfunction in Alzheimer's disease with cerebrovascular disease. *Mol Brain*. 2016;9:27.
- Jayaraman A, Carroll JC, Morgan TE, Lin S, Zhao L, Arimoto JM, et al. 17 β -estradiol and progesterone regulate expression of β -amyloid clearance factors in primary neuron cultures and female rat brain. *Endocrinology*. 2012;153(11):5467–79.
- Nelson LH, Saulsbery AI, Lenz KM. Small cells with big implications: Microglia and sex differences in brain development, plasticity and behavioral health. *Prog Neurobiol*. 2019;176:103–19.
- Hohman TJ, Dumitrescu L, Barnes LL, Thambisetty M, Beecham G, Kunkle B, et al. Sex-Specific Association of Apolipoprotein E With Cerebrospinal Fluid Levels of Tau. *JAMA Neurol*. 2018;75(8):989–98.
- Deming Y, Dumitrescu L, Barnes LL, Thambisetty M, Kunkle B, Gifford KA, et al. Sex-specific genetic predictors of Alzheimer's disease biomarkers. *Acta Neuropathol*. 2018;136(6):857–72.

24. Köglberger S, Cordero-Maldonado ML, Antony P, Forster JJ, Garcia P, Buttini M, et al. Gender-Specific Expression of Ubiquitin-Specific Peptidase 9 Modulates Tau Expression and Phosphorylation: Possible Implications for Tauopathies. *Mol Neurobiol*. 2017;54(10):7979–93.
25. Yu W, Yu W, Yang Y, Lü Y. Exploring the Key Genes and Identification of Potential Diagnosis Biomarkers in Alzheimer's Disease Using Bioinformatics Analysis. *Front Aging Neurosci*. 2021;13:602781.
26. Zhang T, Liu N, Wei W, Zhang Z, Li H. Integrated Analysis of Weighted Gene Coexpression Network Analysis Identifying Six Genes as Novel Biomarkers for Alzheimer's Disease. *Oxid Med Cell Longev*. 2021;2021:9918498.
27. Sood S, Gallagher IJ, Lunnon K, Rullman E, Keohane A, Crossland H, et al. A novel multi-tissue RNA diagnostic of healthy ageing relates to cognitive health status. *Genome Biol*. 2015;16(1):185.
28. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS*. 2012;16(5):284–7.
29. Walter W, Sánchez-Cabo F, Ricote M. GPlot: an R package for visually combining expression data with functional analysis. *Bioinformatics*. 2015;31(17):2912–4.
30. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*. 2005;102(43):15545–50.
31. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res*. 2019;47(D1):D607–13.
32. Yi L, Wu G, Guo L, Zou X, Huang P. Comprehensive Analysis of the PD-L1 and Immune Infiltrates of m6A RNA Methylation Regulators in Head and Neck Squamous Cell Carcinoma. *Mol Ther Nucleic Acids*. 2020;21:299–314.
33. Zeng D, Li M, Zhou R, Zhang J, Sun H, Shi M, et al. Tumor Microenvironment Characterization in Gastric Cancer Identifies Prognostic and Immunotherapeutically Relevant Gene Signatures. *Cancer Immunol Res*. 2019;7(5):737–50.
34. Wang J, Sun J, Liu LN, Flies DB, Nie X, Toki M, et al. Siglec-15 as an immune suppressor and potential target for normalization cancer immunotherapy. *Nat Med*. 2019;25(4):656–66.
35. Wang X, Wang D, Su F, Li C, Chen M. Immune abnormalities and differential gene expression in the hippocampus and peripheral blood of patients with Alzheimer's disease. *Ann Transl Med*. 2022;10(2):29.
36. Park YH, Pyun JM, Hodges A, Jang JW, Bice PJ, Kim S, et al. Dysregulated expression levels of APOB in peripheral blood are associated with brain atrophy and amyloid- β deposition in Alzheimer's disease. *Alzheimers Res Ther*. 2021;13(1):183.
37. Siedlecki-Wullich D, Català-Solsona J, Fábregas C, Hernández I, Clarimon J, Lleó A, et al. Altered microRNAs related to synaptic function as potential plasma biomarkers for Alzheimer's disease. *Alzheimers Res Ther*. 2019;11(1):46.
38. Song YJ, Li SR, Li XW, Chen X, Wei ZX, Liu QS, et al. The Effect of Estrogen Replacement Therapy on Alzheimer's Disease and Parkinson's Disease in Postmenopausal Women: A Meta-Analysis. *Front Neurosci*. 2020;14:157.
39. Uddin MS, Rahman MM, Jakaria M, Rahman MS, Hossain MS, Islam A, et al. Estrogen Signaling in Alzheimer's Disease: Molecular Insights and Therapeutic Targets for Alzheimer's Dementia. *Mol Neurobiol*. 2020;57(6):2654–70.
40. Nazarian A, Yashin AI, Kulminski AM. Genome-wide analysis of genetic predisposition to Alzheimer's disease and related sex disparities. *Alzheimers Res Ther*. 2019;11(1):5.
41. Ding Q, Markesbery WR, Chen Q, Li F, Keller JN. Ribosome dysfunction is an early event in Alzheimer's disease. *J Neurosci*. 2005;25(40):9171–5.
42. Metaxas A, Thygesen C, Kempf SJ, Anzalone M, Vaitheeswaran R, Petersen S, et al. Ageing and amyloidosis underlie the molecular and pathological alterations of tau in a mouse model of familial Alzheimer's disease. *Sci Rep*. 2019;9(1):15758.
43. Santos-Galindo M, Acaz-Fonseca E, Bellini MJ, Garcia-Segura LM. Sex differences in the inflammatory response of primary astrocytes to lipopolysaccharide. *Biol Sex Differ*. 2011;2:7.
44. Klein SL, Morgan R. The impact of sex and gender on immunotherapy outcomes. *Biol Sex Differ*. 2020;11(1):24.
45. Capone I, Marchetti P, Ascierto PA, Malorni W, Gabriele L. Sexual Dimorphism of Immune Responses: A New Perspective in Cancer Immunotherapy. *Front Immunol*. 2018;9:552.
46. Scott SC, Shao XM, Niknafs N, Balan A, Pereira G, Marrone KA, et al. Sex-specific differences in immunogenomic features of response to immune checkpoint blockade. *Front Oncol*. 2022;12:945798.
47. Busse S, von Hoff F, Michler E, Hartig R, Bogerts B, Busse M. Altered expression of costimulatory molecules in dementias. *Eur Arch Psychiatry Clin Neurosci*. 2022;272(5):807–815.
48. Holmes S, Singh M, Su C, Cunningham RL. Effects of Oxidative Stress and Testosterone on Pro-Inflammatory Signaling in a Female Rat Dopaminergic Neuronal Cell Line. *Endocrinology*. 2016;157(7):2824–35.
49. Snyder B, Duong P, Trieu J, Cunningham RL. Androgens modulate chronic intermittent hypoxia effects on brain and behavior. *Horm Behav*. 2018;106:62–73.
50. Budoff MJ, Ellenberg SS, Lewis CE, Mohler ER 3rd, Wenger NK, Bhasin S, et al. Testosterone Treatment and Coronary Artery Plaque Volume in Older Men With Low Testosterone. *JAMA*. 2017;317(7):708–16.
51. Tenkorang MAA, Duong P, Cunningham RL. NADPH Oxidase Mediates Membrane Androgen Receptor-Induced Neurodegeneration. *Endocrinology*. 2019;160(4):947–63.
52. Tenkorang MA, Snyder B, Cunningham RL. Sex-related differences in oxidative stress and neurodegeneration. *Steroids*. 2018;133:21–7.
53. Duong P, Tenkorang MAA, Trieu J, McCuiston C, Rybalchenko N, Cunningham RL. Neuroprotective and neurotoxic outcomes of androgens and estrogens in an oxidative stress environment. *Biol Sex Differ*. 2020;11(1):12.
54. Cunningham RL, Singh M, O'Bryant SE, Hall JR, Barber RC. Oxidative stress, testosterone, and cognition among Caucasian and Mexican-American men with and without Alzheimer's disease. *J Alzheimers Dis*. 2014;40(3):563–73.
55. Beamish SB, Frick KM. A Putative Role for Ubiquitin-Proteasome Signaling in Estrogenic Memory Regulation. *Front Behav Neurosci*. 2022;15:807215.
56. Hales CM, Dammer EB, Diner I, Yi H, Seyfried NT, Gearing M, et al. Aggregates of small nuclear ribonucleic acids (snRNAs) in Alzheimer's disease. *Brain Pathol*. 2014;24(4):344–51.
57. Hsieh YC, Guo C, Yalamanchili HK, Abreha M, Al-Ouran R, Li Y, et al. Tau-Mediated Disruption of the Spliceosome Triggers Cryptic RNA Splicing and Neurodegeneration in Alzheimer's Disease. *Cell Rep*. 2019;29(2):301–316.e10.
58. Cheng Z, Shang Y, Xu X, Dong Z, Zhang Y, Du Z, et al. Presenilin 1 mutation likely contributes to U1 small nuclear RNA dysregulation and Alzheimer's disease-like symptoms. *Neurobiol Aging*. 2021;100:1–10.
59. Ragan C, Patel K, Edson J, Zhang ZH, Gratten J, Mowry B. Small non-coding RNA expression from anterior cingulate cortex in schizophrenia shows sex specific regulation. *Schizophr Res*. 2017;183:82–7.
60. Liu P, Liu Z, Wang J, Ma X, Dang Y. HINT1 in Neuropsychiatric Diseases: A Potential Neuroplastic Mediator. *Neural Plast*. 2017;2017:5181925.
61. Chu L, Fu G, Meng Q, Zhou H, Zhang M. Identification of urinary biomarkers for type 2 diabetes using bead-based proteomic approach. *Diabetes Res Clin Pract*. 2013;101(2):187–93.
62. Yu W, Yin H, Sun Y, Shi S, Li J, Wang X. The attenuation effect of potassium 2-(1-hydroxypentyl)-benzoate in a mouse model of diabetes-associated cognitive decline: The protein expression in the brain. *CNS Neurosci Ther*. 2022;28(7):1108–23.
63. Chen Q, Wang X, O'Neill FA, Walsh D, Kendler KS, Chen X. Is the histidine triad nucleotide-binding protein 1 (HINT1) gene a candidate for schizophrenia? *Schizophr Res*. 2008;106(2–3):200–7.
64. Tao Y, Han Y, Yu L, Wang Q, Leng SX, Zhang H. The Predicted Key Molecules, Functions, and Pathways That Bridge Mild Cognitive Impairment (MCI) and Alzheimer's Disease (AD). *Front Neurol*. 2020;11:233.
65. Mishra A, Wang Y, Yin F, Vitali F, Rodgers KE, Soto M, et al. A tale of two systems: Lessons learned from female mid-life aging with implications for Alzheimer's prevention & treatment. *Ageing Res Rev*. 2022;74:101542.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.