Identification of diagnostic biomarkers in Alzheimer's disease by integrated bioinformatic analysis and machine learning strategies

Boru Jin[1,2†], Xiaoqin Cheng[1,2†], Guoqiang Fei[1,2], Shaoming Sang[3]* and Chunjiu Zhong[1,2]*

[1]Department of Neurology, Zhongshan Hospital, Fudan University, Shanghai, China, [2]Collaborative Innovation Center for Brain Science, Fudan University, Shanghai, China, [3]Shanghai Raising Pharmaceutical Technology Co., Ltd. Shanghai, China

Background: Alzheimer's disease (AD) is the most prevalent form of dementia, and is becoming one of the most burdening and lethal diseases. More useful biomarkers for diagnosing AD and reflecting the disease progression are in need and of significance.

Methods: The integrated bioinformatic analysis combined with machine-learning strategies was applied for exploring crucial functional pathways and identifying diagnostic biomarkers of AD. Four datasets (GSE5281, GSE131617, GSE48350, and GSE84422) with samples of AD frontal cortex are integrated as experimental datasets, and another two datasets (GSE33000 and GSE44772) with samples of AD frontal cortex were used to perform validation analyses. Functional Correlation enrichment analyses were conducted based on Gene ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG), and the Reactome database to reveal AD-associated biological functions and key pathways. Four models were employed to screen the potential diagnostic biomarkers, including one bioinformatic analysis of Weighted gene co-expression network analysis (WGCNA)and three machine-learning algorithms: Least absolute shrinkage and selection operator (LASSO), support vector machine-recursive feature elimination (SVM-RFE) and random forest (RF) analysis. The correlation analysis was performed to explore the correlation between the identified biomarkers with CDR scores and Braak staging.

Results: The pathways of the immune response and oxidative stress were identified as playing a crucial role during AD. Thioredoxin interacting protein (TXNIP), early growth response 1 (EGR1), and insulin-like growth factor binding protein 5 (IGFBP5) were screened as diagnostic markers of AD. The diagnostic efficacy of TXNIP, EGR1, and IGFBP5 was validated with corresponding AUCs of 0.857, 0.888, and 0.856 in dataset GSE33000, 0.867, 0.909, and 0.841 in dataset GSE44770. And the AUCs of the combination of these three biomarkers as a diagnostic tool for AD were 0.954 and 0.938 in the two verification datasets.

Conclusion: The pathways of immune response and oxidative stress can play a crucial role in the pathogenesis of AD. TXNIP, EGR1, and IGFBP5 are useful biomarkers for diagnosing AD and their mRNA level may reflect the development of the disease by correlation with the CDR scores and Breaking staging.

Introduction

Alzheimer's disease (AD) is the most prevalent form of dementia accounting for 60–80% of all cases (Alzheimer's Association, 2022), and is becoming one of the main causes of death and posing a huge burden on patients and their families (Katzman, 2008; Georges et al., 2020). It is reported that about 50% of people aged 80 suffer from this disorder (Nalbantoglu et al., 2005) and the number of that would accumulate up to 115 million by 2050, which means there are 7.7 million increased cases every year and one more suffers every 4 s (Sosa-Ortiz et al., 2012). AD is manifested by memory loss, executive dysfunctions, and other cognitive deficits affecting patients' ability to perform everyday activities (Querfurth and Laferla, 2010; Mckhann et al., 2011) and would eventually lead to the premature death of an individual occurring typically 3–9 years after diagnosis (Querfurth and Laferla, 2010). Due to the lack of effective treatments and the increasing average lifespan, AD has posed an enormous burden on worldwide economics and health (Alzheimer's Association, 2016; Robinson et al., 2017).

The major neuropathological features of AD are intracellular neurofibrillary tangles (NFTs) formed by hyperphosphorylated tau protein, extracellular senile plaques composed of aggregated β-amyloid (Aβ) fibers (Tabaton et al., 1991; Smith, 1998), and progressive brain atrophy causing by loss of synapses and neurons (Butterfield and Halliwell, 2019; Alzheimer's Association, 2021). Recently, attention has also been paid to other pathological markers including insulin resistance (de la Monte, 2017; Rad et al., 2018), oxidative stress (Perry et al., 2008; Jiang et al., 2016), neuroinflammation (Calsolaro and Edison, 2016), erythrocytic abnormality (Kosenko et al., 2020), mitochondrial dysfunction (Lustbader, 2004; Carvalho et al., 2019), and so forth. Several novel hypotheses were proposed such as the erythrocytic hypothesis (Kosenko et al., 2020), heart failure link to AD (Tublin et al., 2019), synaptic failure hypothesis (Tublin et al., 2019), and mitochondrial cascade hypothesis (Swerdlow et al., 2014). However, lacking a comprehensive understanding of the whole mechanism, none of these could precisely connect all the pathological events. There is an urgent need to detangle the mechanism of AD and identify useful biomarkers for diagnosis.

Bioinformatics analysis has evolved into an integrative field between computer science and biology, which allows the representation, storage, management, analysis, and investigation of numerous data types with diverse algorithms and computational tools (Mulder et al., 2017; Auslander et al., 2021). However, due to the quick development of next-generation sequencing and other emerging omics techniques, accumulated omics data at an astonishing speed and scope is urging for more effective approaches to conduct sophisticated analyses from various biomolecular levels, such as genomics, transcriptomics, proteomics, radiomics and metabolomics (Pevsner, 2015; Ayyildiz and Piazza, 2019). Fortunately, machine learning meets omics and exhibits extreme power in processing and modeling omics data with huge and diverse volumes (Li et al., 2022). Machine learning is a branch of artificial intelligence focusing on simulating human learning by exploring patterns in the data and applying self-improvement to continually enhance the performance of learning tasks (Auslander et al., 2021). Recently, the integrated bioinformatic analysis combined with machine-learning strategies was applied to the identification of potential pathways and diagnostic biomarkers of diseases, which has earned some praise (Journal of Nature Genetics, 2019; Auwul et al., 2021; Tran et al., 2021).

In our study, we integrated four frontal cortical datasets from the GEO database to discover novel pathways and identify diagnostic biomarkers of AD by bioinformatic analysis combined with machine learning strategies. The differential expressions and diagnostic efficacy of the identified biomarkers were verified in another two frontal cortical datasets of AD. The correlation analysis between the identified biomarkers and the CDR scores and Braaking staging. Finally, biomarkers associated with the key functional pathways in AD were identified and verified, which could also reflect the development of AD.

Materials and methods

A diagram of the workflow of the bioinformatics analyses combined with machine learning strategies is shown in **Figure 1**.

Data collection and data processing

We retrieved and downloaded six microarray expression profile datasets with the frontal cortex of AD patients from the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (GEO) database.[1] The search was conducted with the following keywords: ("Alzheimer's disease" and "Expression profiling by array"), and the species was restricted as "Homo sapiens." Four datasets (GSE5281, GSE131617, GSE48350, and GSE84422) from the platform of Affymetrix are used as

---

1 http://www.ncbi.nlm.nih.gov/geo

FIGURE 1
The workflow of the analysis process.
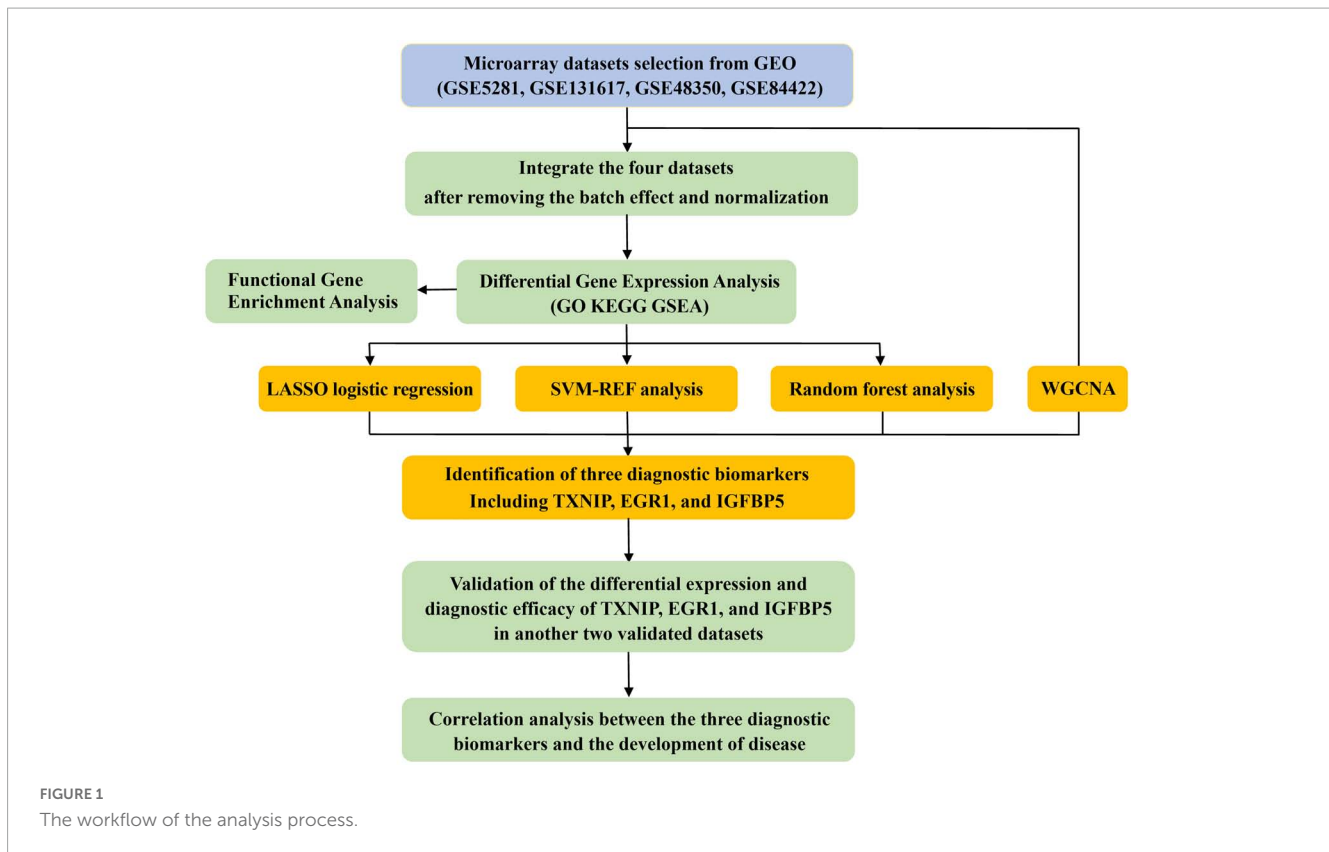
experimental datasets, and another two datasets (GSE33000 and GSE44772) from the platform of Rosetta/Merck were used as validation datasets.

Initially, referring to the annotation from the platforms, probes were annotated with gene symbols. Once there are multiple probes associated with the same gene, the average value of the expression would be calculated and applied. Then, the batch effect of the four experimental datasets was removed using the "Combat" function of the "SVA" package in R to fulfill normalization (Johnson et al., 2007). Finally, the validation of the differential expressions and diagnostic efficacy of identified biomarkers was performed on the validation datasets.

Differential gene expression analysis

After normalization, four datasets were merged into an integrated dataset including 87 frontal cortical samples of AD and 126 controls. The differential gene expression analysis was conducted on this integrated dataset with the "limma" package in R. The $|\log2FC|$ (fold change) >2 and adjusted $p < 0.05$ were regarded as thresholds for the screening. Heatmaps and volcano plots were performed with "pheatmap" and "EnhancedVolcano" packages in R.
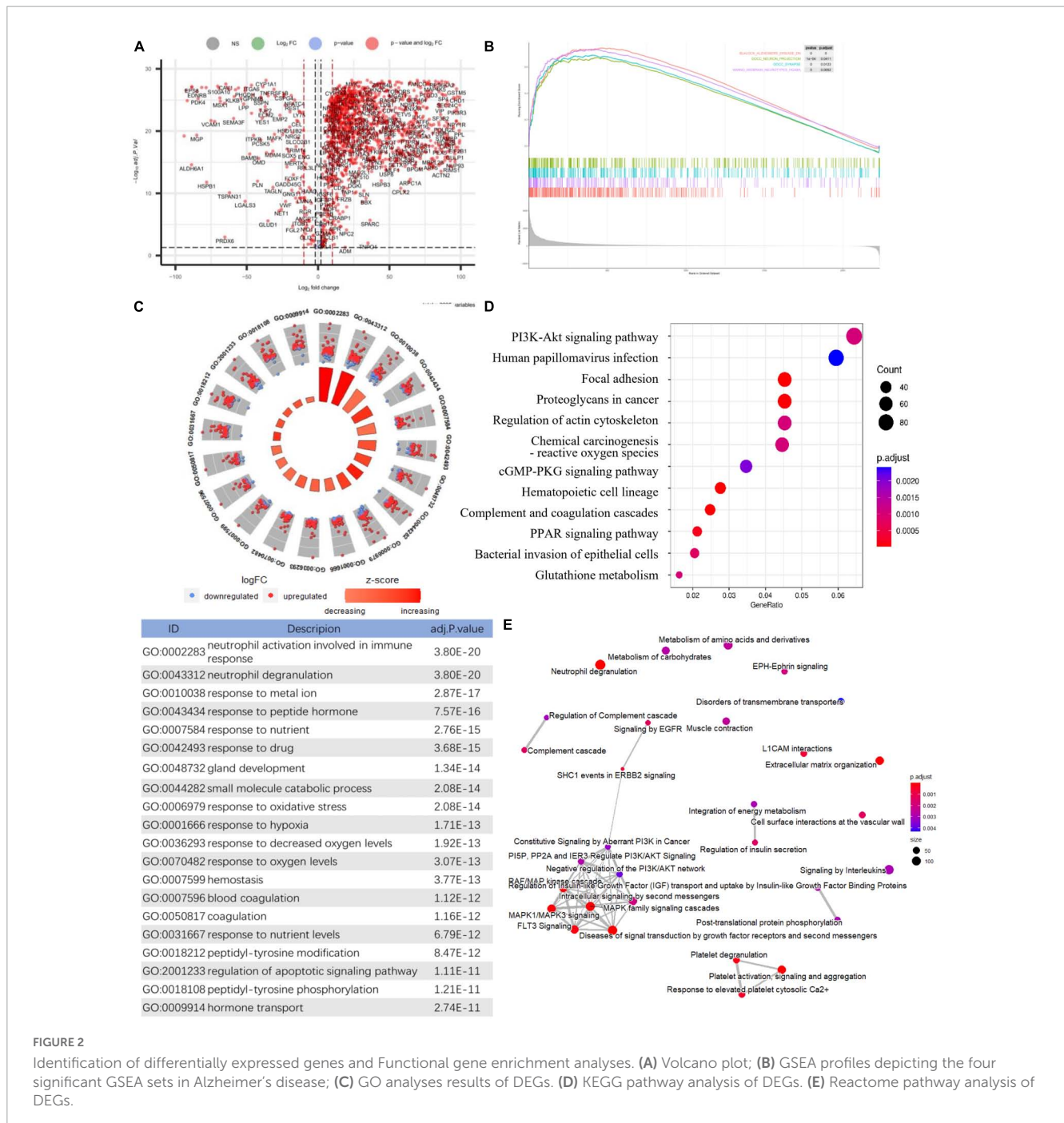
Functional enrichment analysis

Focusing on all genes instead of only DEGs and demonstrating significantly enriched functional pathways more intuitively, gene

set enrichment analysis (GSEA) was performed in R with clusterProfiler (Subramanian et al., 2005; Yu et al., 2012). Gene ontology (GO) enrichment analysis was conducted considering three hierarchical categories of biological process, molecular function, and cellular component with the "clusterProfiler" package in R (Yu et al., 2012). Pathway enrichment analysis was performed on Kyoto Encyclopedia of Genes and Genomes (KEGG) and Reactome database with "clusterProfiler" and "ReactomePA" packages in R (Yu et al., 2012; Yu and He, 2016; Minoru et al., 2017; Bijay et al., 2019).

Screening the diagnostic biomarkers

Four models were applied to screen the potential diagnostic biomarkers, including one bioinformatic analysis and three machine-learning algorithms. Weighted gene co-expression network analysis (WGCNA) is a bioinformatic method describing the correlation between genes and sample traits, which has been widely used for identifying candidate biomarkers or therapeutic targets (Langfelder et al., 2009). The least absolute shrinkage and selection operator (LASSO) is a shrinkage and variable selection method for regression models, which was applied to identify the diagnostic genes associated with discrimination with the "glmnet" package in R (Ranstam and Cook, 2018). Support vector machine-recursive feature elimination (SVM-RFE) was conducted in R using the "e1071" package with fivefold cross-validation (Sanz et al., 2018). Random forest (RF) analysis was performed in R with the "randomForest" package (Rigatti, 2017). Finally, the Venn diagram was plotted to visualize the overlapping potential

FIGURE 2
Identification of differentially expressed genes and Functional gene enrichment analyses. **(A)** Volcano plot; **(B)** GSEA profiles depicting the four significant GSEA sets in Alzheimer's disease; **(C)** GO analyses results of DEGs. **(D)** KEGG pathway analysis of DEGs. **(E)** Reactome pathway analysis of DEGs.

biomarkers among the four models as the candidate biomarkers (Bardou et al., 2014).
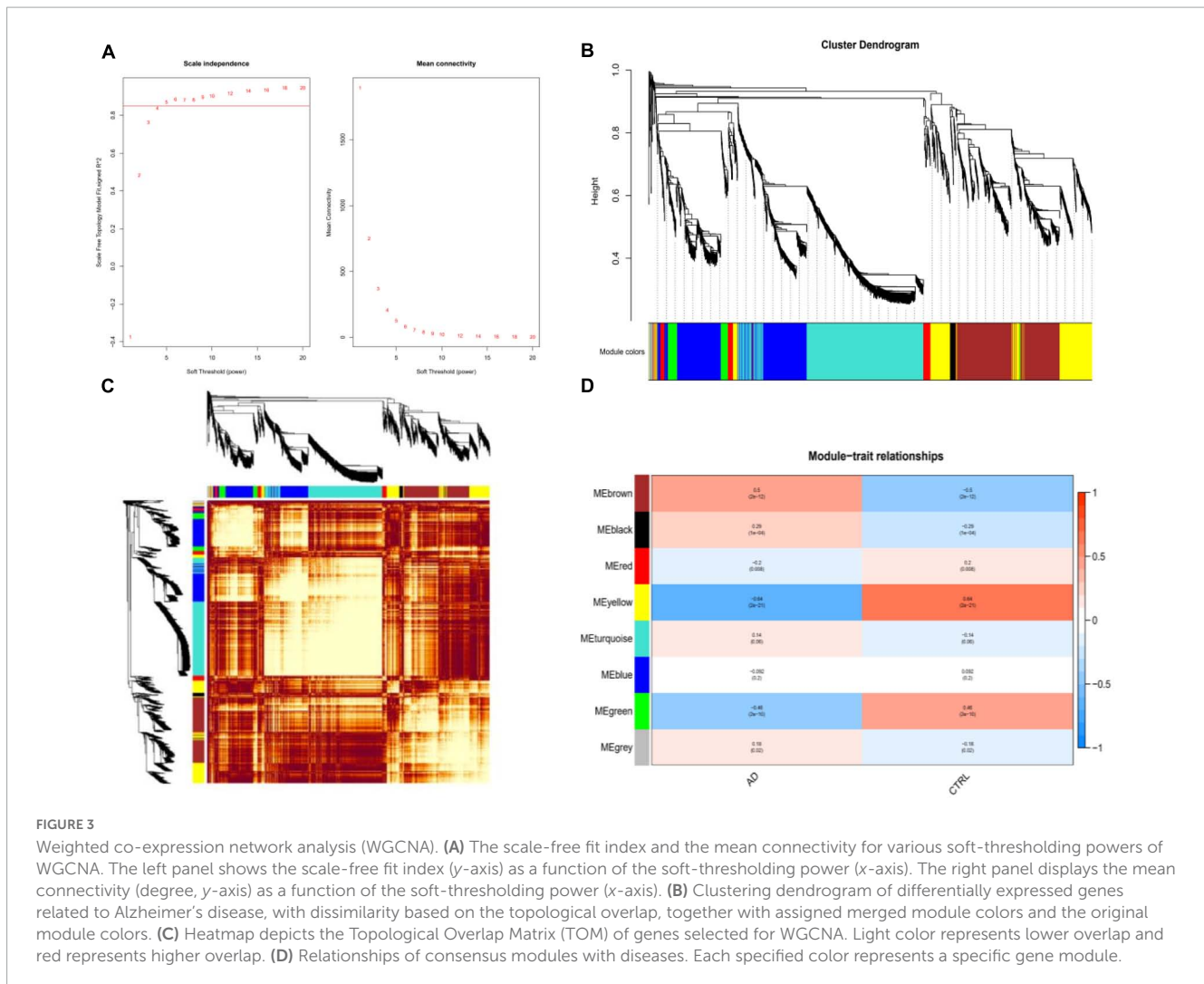
Validation of the candidate biomarkers

The differential expression of the candidate biomarkers was verified in the validation datasets of GSE33000 and GSE44772. The diagnostic efficacy of the candidate biomarkers was evaluated by receiver operating characteristic (ROCs) analysis with the area under the curves (AUCs) (Seshan et al., 2013). The correlation analysis was performed to explore the correlation between the candidate biomarkers with CDR scores and Braak staging.

Results

Identification of differentially expressed genes

After normalization, an integrated dataset with frontal cortical samples of AD was formed by four GEO datasets (GSE5281, GSE131617, GSE48350, and GSE84422), consisting of 87 AD patients and 126 control subjects. In the integrated dataset, the differential gene expression analysis identified 2235 DEGs, including 2029 downregulated genes and 206 upregulated genes in AD compared to the matched controls (**Figure 2A** and

FIGURE 3

Weighted co-expression network analysis (WGCNA). **(A)** The scale-free fit index and the mean connectivity for various soft-thresholding powers of WGCNA. The left panel shows the scale-free fit index (y-axis) as a function of the soft-thresholding power (x-axis). The right panel displays the mean connectivity (degree, y-axis) as a function of the soft-thresholding power (x-axis). **(B)** Clustering dendrogram of differentially expressed genes related to Alzheimer's disease, with dissimilarity based on the topological overlap, together with assigned merged module colors and the original module colors. **(C)** Heatmap depicts the Topological Overlap Matrix (TOM) of genes selected for WGCNA. Light color represents lower overlap and red represents higher overlap. **(D)** Relationships of consensus modules with diseases. Each specified color represents a specific gene module.

Supplementary Table 1). The visualized DEG expressions in the integrated dataset were shown in the heatmap (Supplementary Figure 1).

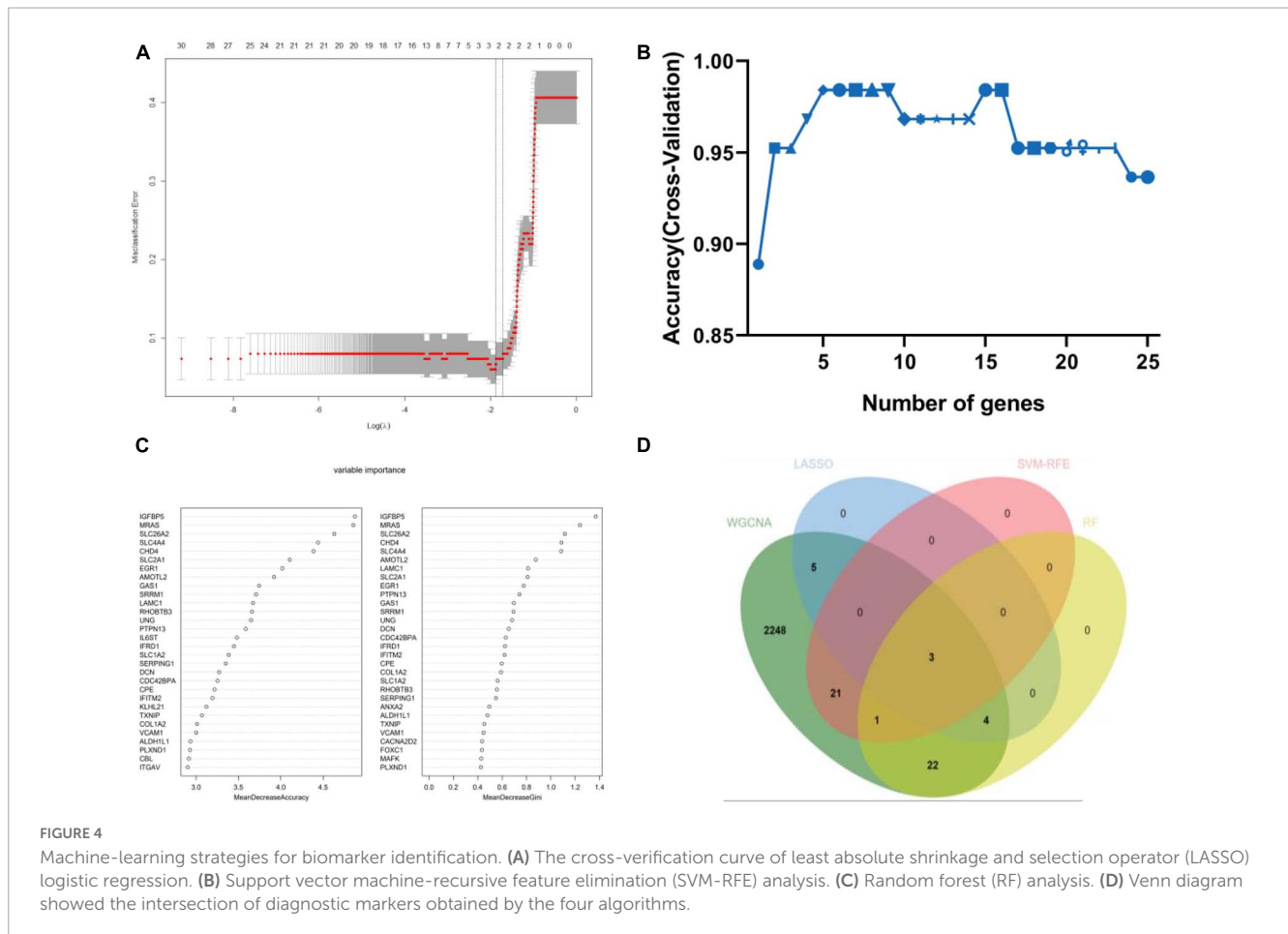## Functional gene enrichment analyses

Gene set enrichment analysis in all detectable genes showed that genes in AD were mainly enriched in the following pathways: AD, neuron projection, synapses, multiple midbrain neurotypes, and so forth (Figure 2B). Also, GSEA_GO analysis revealed consistent results in pathway identification with GSEA analysis, in which dendrite and glutamatergic synapses were involved (Supplementary Table 1). GSEA_KEGG analysis showed that neurodegeneration and oxygen-related pathways were involved in AD, including the HIF-1 signaling and oxidative phosphorylation pathways (Supplementary Table 1). The results of the above GSEA analyses were further confirmed by GO, KEGG, and Reactome analysis on DEGs (Supplementary Table 1). The biological processes associated with immune response, oxidative stress, and apoptosis were significantly enriched by GO analysis (Figure 2C). PI3K-Akt signaling pathway was highlighted in both KEGG and Reactome analysis (Figures 2D, E).

Screening the diagnostic biomarkers

Weighted gene co-expression network analysis showed that eight remarkable co-expression gene modules identified were significantly correlated with AD (Figures 3A–D). LASSO logistic regression algorithm screened twelve potential diagnostic markers from DEGs (Figure 4A). SVM-RFE and RF analyses showed that there were 25 and 30 potential diagnostic markers associated with AD (Figures 4B, C). Among these potential biomarkers, there were three overlapping genes: thioredoxin interacting protein (TXNIP), early growth response 1 (EGR1), and insulin-like growth factor binding protein 5 (IGFBP5) (Figure 4D and Supplementary Table 1).

Validation of the candidate biomarkers

The expression changes of TXNIP, EGR1, and IGFBP5 were further validated in another two datasets GSE33000 and GSE44770. The results were consistent with the integrated dataset, in which TXNIP and IGFBP5 were significantly upregulated and EGR1 was downregulated (Figures 5A, B). The ROC analysis showed that the

FIGURE 4

Machine-learning strategies for biomarker identification. **(A)** The cross-verification curve of least absolute shrinkage and selection operator (LASSO) logistic regression. **(B)** Support vector machine-recursive feature elimination (SVM-RFE) analysis. **(C)** Random forest (RF) analysis. **(D)** Venn diagram showed the intersection of diagnostic markers obtained by the four algorithms.

AUCs of TXNIP, EGR1, and IGFBP5 were 0.857, 0.888, and 0.856 in dataset GSE33000, 0.867, 0.909, and 0.841 in dataset GSE44770. The AUCs of the combination of these three biomarkers as a diagnostic tool for AD were 0.954 and 0.938 (Figure 5C). The correlation analysis indicated that TXNIP and IGFBP5 expressions were significantly correlated with CDR scores, and EGR1 and IGFBP5 expressions were significantly correlated with the Braak staging (Figure 6).

Discussion

The frontal cortex has always been viewed as the "motor" lobe associated with two cognitive functions of memory and motor (Fuster, 1993; Boyle, 2004; Hashimoto et al., 2017) and is very vulnerable to suffering from impairment in AD. And studies suggested that the frontal cortex is quite sensitive to subclinical changes which may help predict cognitive impairments and early disturbance of daily activities (Stoeckel et al., 2013; Marshall et al., 2019). Hence, we select the frontal cortical samples of AD to identify the diagnostic biomarkers. To enlarge the scale of the sample size and simultaneously avoid the batch effect, we integrated four AD frontal cortical datasets of transcriptome from the same platform to conduct the analyses, and another two frontal cortical datasets to perform further validation. Considering the extreme power of machine learning in processing and modeling omics data,

the integrated bioinformatic analysis was combined with three machine-learning strategies to fulfill the identification of diagnostic biomarkers.

The differential gene expression analysis revealed 2235 DEGs in AD compared to the matched controls, including 2029 downregulated genes and 206 upregulated genes (Figure 2A). Multiple functional enrichment analyses found that several essential pathways were significantly enriched in AD patients, including AD, neurodegeneration, synapse, immune response, oxidative stress, apoptotic signaling pathway, and so forth. It is widely known that the loss of synapse correlates the best with cognitive impairment and even precedes neuronal loss in AD, and there are many factors contributing to synaptic dysfunction in AD, especially the above-identified: immune response and oxidative stress (DeKosky and Scheff, 1990; Britschgi and Wyss-Coray, 2007; Hong et al., 2016). Mounting studies show that oxidative stress can impair synapses and contribute to AD through most pathological hypotheses including the amyloid cascade hypothesis, tau hypothesis, inflammatory hypothesis, and so forth (Ansari and Scheff, 2010; Zhao and Zhao, 2013; Bai et al., 2022). Accumulating evidence has also stressed that immune responses involving glial cells and the complement system are prominently activated in the AD brain, which can prune excess synapses inappropriately and mediate synapse loss eventually (Akiyama et al., 2000; Wyss-Coray, 2006; Britschgi and Wyss-Coray, 2007; Hong et al., 2016; Rajendran and Paolicelli, 2018). Generally, our results confirmed
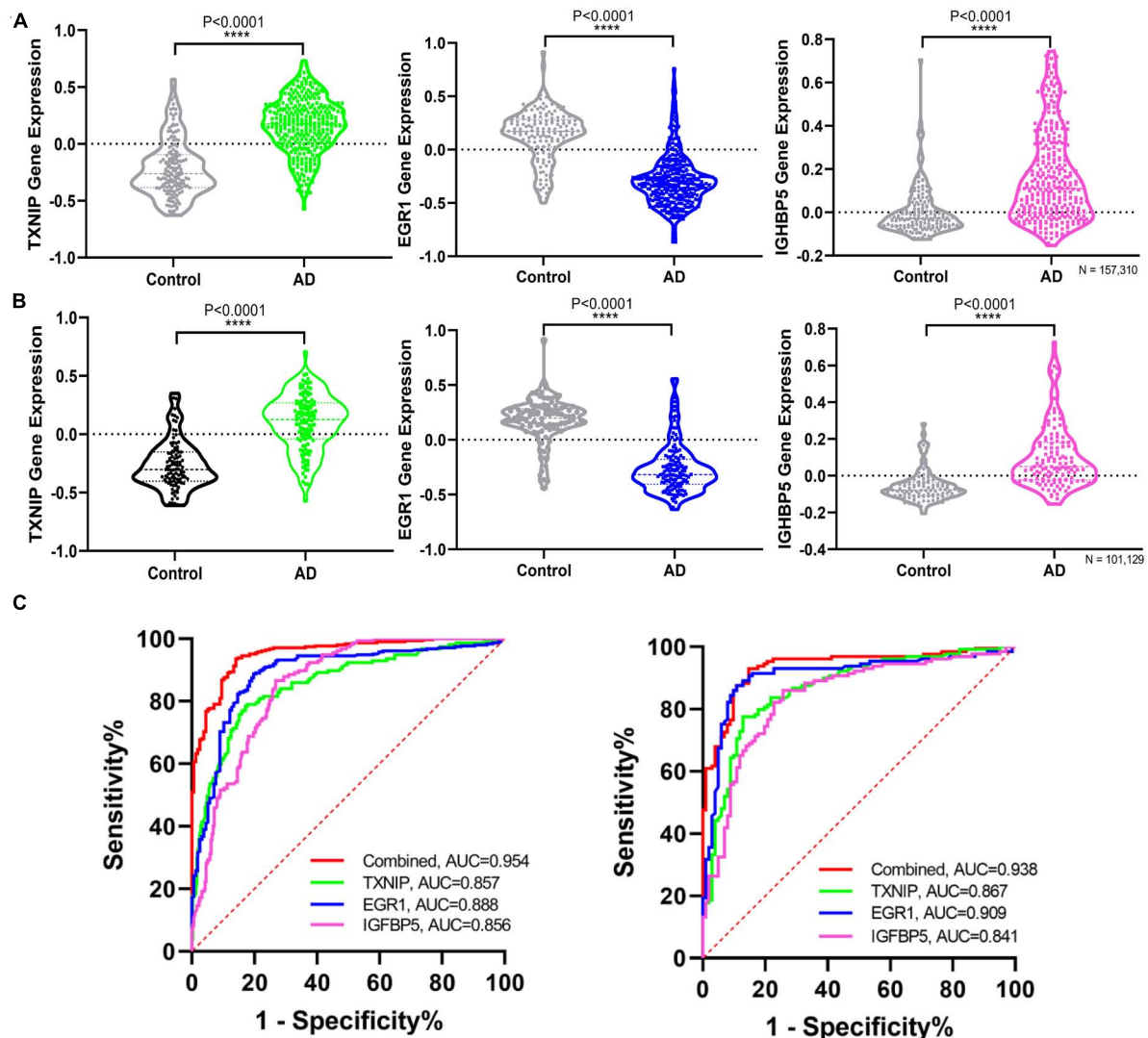
FIGURE 5
Validating the differential expression and diagnostic efficacy of the identified biomarkers. **(A)** Validation of the expression levels of the identified biomarkers in dataset GSE33000. **(B)** Validation of the expression levels of the identified biomarkers in dataset GSE44770. **(C)** Validation of the diagnostic efficacy of diagnostic biomarkers revealed by ROC analysis in the validation dataset GSE33000 and GSE44770 (TXNIP, EGR1, IGFBP5, and the combination of the three genes as a diagnostic tool).

the pathological pathways of synapse and apoptosis in AD and further stressed the crucial role of the immune response and oxidative stress in the pathogenesis of the disease.

Moreover, one bioinformatic analysis of WGCNA and three machine-learning strategies of LASSO, SVM-RFE, and RF analyses commonly identified that TXNIP, EGR1, and IGFBP5 could serve as biomarkers of AD, combining them as a tool gave rise to high AUCs of 0.954 and 0.938 in the two verification datasets (**Figure 5C**). The correlation analysis further revealed that the expressions of TXNIP and IGFBP5 were significantly correlated with the CDR scores, and the expressions of EGR1 and IGFBP5 were significantly correlated with the Braak staging (**Figure 6**).

Previous studies have shown that TXNIP as an endogenous inhibitor of antioxidant thioredoxin was found to increase in AD patients and AD mouse models, and could be a key coordinator of different pathological processes (Tsubaki et al., 2020). TXNIP connects oxidative stress and inflammation by interaction with the nucleotide-binding domain, leucine-rich-containing family, and pyrin domain-containing-3 (NLRP3) inflammasome complex (Wang et al., 2019; Eraky and Ramadan, 2022; Sbai et al., 2022). Recent studies also suggested that blocking the interaction of NLRP3 provides a significant effect, and thus TXNIP could serve as a therapeutic target (Zhang et al., 2021). Therefore, TXNIP closely associated with the identified pathways of immune response and oxidative stress can be a useful biomarker of AD. EGR1 has also been reported in previous gene-wide association analyses using brain expression data (Koldamova et al., 2014; Mukherjee et al., 2017; Lim et al., 2018), which was associated with Aβ toxicity and was invalided in a C. elegans model (Mukherjee et al., 2017). Functionally, EGR1 helps maintain the brain's cholinergic function during AD by regulating acetylcholinesterase (AChE) (Hu et al., 2019). EGR1 can bind to the BACE1 promoter and block the
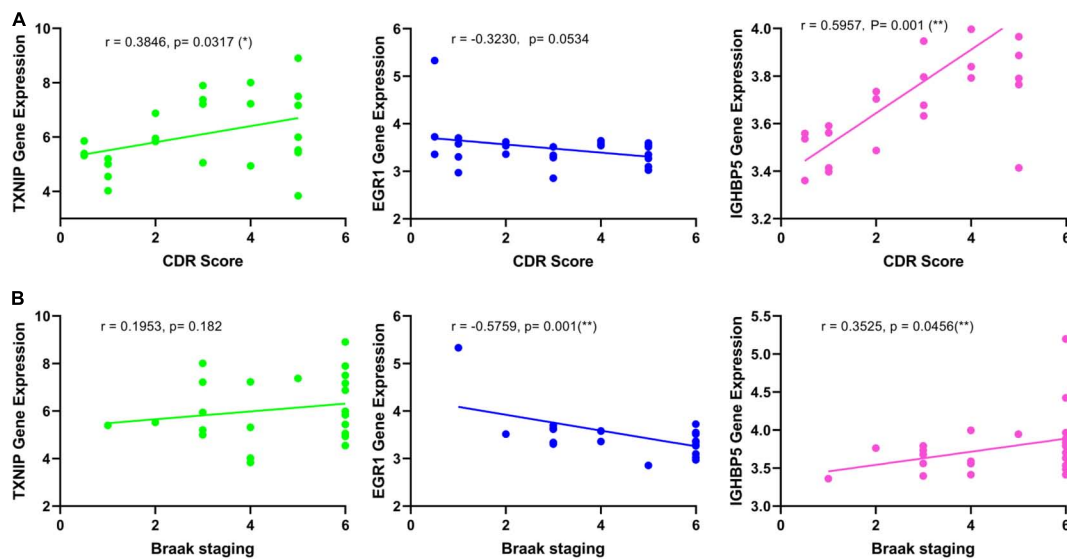
FIGURE 6
Correlation analysis of the identified biomarkers with the development of AD. **(A)** Correlation of the identified biomarkers with the CDR scores. **(B)** Correlation of the identified biomarkers with the Braak staging.

activation of the APP signaling to ultimately suppress the Aβ deposition and improve the cognitive function of AD (He et al., 2022). Studies also proposed EGR1 to be a key molecule affecting the activity of the nucleus basalis of Meynert by regulating synaptic activity and plasticity during AD (Zhu et al., 2016). Conclusively, EGR1 plays an important role in the development of AD and can serve as a useful biomarker. IGFBP5 is a pluripotent growth factor supporting neuronal survival and axon growth (Caroni and Grandes, 1990; Bach et al., 2005; Fernandez and Torres-Alemán, 2012; Rauskolb et al., 2017), which can coordinate the bioavailability and bioactivity of insulin-like growth factor 1 (IGF-1). IGFBP5 can modulate lipid metabolism and insulin sensitivity (Xiao et al., 2020), which are both associated with the cognitive impairment (Kao et al., 2020; Kellar and Craft, 2020). Studies have shown that IGFBP5 was associated with faster cognitive decline (Yu et al., 2018; Kim et al., 2019) and was found to increase in the brains (Rauskolb et al., 2022), cerebrospinal fluid (Salehi et al., 2008), and animal models of AD (Barucker et al., 2015).

Together, TXNIP, EGR1, and IGFBP5 served as potential biomarkers for AD diagnosis reflecting different pathogenetic pathways involved in the development of AD, which may be due to the complicated and multiple pathophysiological manifestations of AD. Besides the EGR1-associated Aβ deposition which has gained the most concern, our results suggested that the TXNIP-associated pathways of the immune response, oxidative stress, and especially their interaction should be paid more attention. More importantly, the identification of IGFBP5 highlights the role of insulin metabolism in the pathogenesis and development of AD. Evidence from epidemiological, clinical, and neuropathology has shown that patients with diabetes are at higher risk of developing AD due to impaired brain insulin signaling (Barbiellini Amidei et al., 2021; De Felice et al., 2022). Studies have repurposed anti-diabetes agents as novel therapeutics for AD, while how impaired insulin signaling and brain insulin resistance occurs remains unclear, urging further exploration.

There are some limitations of our study. Firstly, although we tried to select the same region of frontal cortex in AD brains from the same platform and have performed validation in another two verification datasets, the results still need more experimental confirmation for the data is from publicly available microarray datasets. Secondly, given the limited scale of sample size and type, the diagnostic efficacy of biomarkers should be further explored clinically, and even in samples of blood and cerebrospinal fluid. Thirdly, we fail to identify the early detecting biomarkers of AD due to the lack of datasets on patients with mild cognitive impairment (MCI), though early detection is the now most urgent need under the huge burden of increasing incidence and heavy cost. And we would like to perform the analysis of identifying the potential diagnostic biomarkers of MCI, once there were available datasets.

Conclusion

The integrated bioinformatic analysis combined with machine learning strategies can effectively help identify the functional pathways and diagnostic biomarkers in disease. Based on these methods, we stressed the crucial roles of immune response and oxidative stress in the pathogenesis of AD and identified three genes associated with the above two pathways as useful biomarkers, including TXNIP, EGR1, and IGFBP5. Furthermore, the expressions of TXNIP, EGR1, and IGFBP5 may reflect the development of AD by correlation with the CDR scores and Braak staging.

Data availability statement

The data presented in the study are deposited in the Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo) repository, accession numbers GSE5281, GSE131617, GSE48350, GSE84422, GSE33000, and GSE44772.

Author contributions

Funding

Conflict of interest

CZ, the corresponding author, holds shares of Shanghai Raising Pharmaceutical Co., Ltd., which is dedicated to developing drugs for the prevention and treatment of AD.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnagi.2023.1169620/full#supplementary-material

References

Akiyama, H., Barger, S., Barnum, S., Bradt, B., Bauer, J., Cole, G., et al. (2000). Inflammation and Alzheimer's disease. *Neurobiol. Aging.* 21, 383–421. doi: 10.1016/S0197-4580(00)00124-X

Alzheimer's Association (2016). 2016 Alzheimer's disease facts and figures. *Alzheimer's Dement.* 12, 459–509. doi: 10.1016/j.jalz.2016.03.001

Alzheimer's Association (2021). 2021 Alzheimer's disease facts and figures. *Alzheimers Dement.* 17, 327–406.

Alzheimer's Association (2022). 2022 Alzheimer's disease facts and figures. *Alzheimers Dement.* 18, 700–789. doi: 10.1002/alz.12638

Ansari, M., and Scheff, S. (2010). Oxidative stress in the progression of Alzheimer disease in the frontal cortex. *J. Neuropathol. Exp. Neurol.* 69, 155–167. doi: 10.1097/NEN.0b013e3181cb5af4

Auslander, N., Gussow, A., and Koonin, E. (2021). Incorporating machine learning into established bioinformatics frameworks. *Int. J. Mol. Sci.* 22:2903. doi: 10.3390/ijms22062903

Auwul, M., Rahman, M., Gov, E., Shahjaman, M., and Moni, M. (2021). Bioinformatics and machine learning approach identifies potential drug targets and pathways in COVID-19. *Brief. Bioinform.* 22:bbab120. doi: 10.1093/bib/bbab120

Ayyildiz, D., and Piazza, S. (2019). Introduction to bioinformatics. *Methods Mol. Biol.* 1986, 1–15. doi: 10.1007/978-1-4939-9442-7_1

Bach, L., Headey, S., and Norton, R. (2005). IGF-binding proteins–the pieces are falling into place. *Trends Endocrinol. Metab.* 16, 228–234. doi: 10.1016/j.tem.2005.05.005

Bai, R., Guo, J., Ye, X., Xie, Y., and Xie, T. (2022). Oxidative stress: the core pathogenesis and mechanism of Alzheimer's disease. *Ageing Res. Rev.* 77:101619. doi: 10.1016/j.arr.2022.101619

Barbiellini Amidei, C., Fayosse, A., Dumurgier, J., Machado-Fragua, M., Tabak, A., van Sloten, T., et al. (2021). Association between age at diabetes onset and subsequent risk of dementia. *JAMA* 325, 1640–1649. doi: 10.1001/jama.2021.4001

Bardou, P., Mariette, J., Escudié, F., Djemiel, C., and Klopp, C. (2014). jvenn: an interactive venn diagram viewer. *BMC Bioinformat.* 15:293. doi: 10.1186/1471-2105-15-293

Barucker, C., Sommer, A., Beckmann, G., Eravci, M., Harmeier, A., Schipke, C., et al. (2015). Alzheimer amyloid peptide aβ42 regulates gene expression of transcription and growth factors. *J. Alzheimers Dis.* 44, 613–624. doi: 10.3233/JAD-141902

Bijay, J., Lisa, M., Guilherme, V., Gong, C., Pascual, L., Antonio, F., et al. (2019). The reactome pathway knowledgebase. *Nucleic Acids Res.* 48, D498–D503.

Boyle, P. (2004). Assessing and predicting functional impairment in Alzheimer's disease: the emerging role of frontal system dysfunction. *Curr. Psychiatry Rep.* 6, 20–24. doi: 10.1007/s11920-004-0033-9

Britschgi, M., and Wyss-Coray, T. (2007). Systemic and acquired immune responses in Alzheimer's disease. *Int. Rev. Neurobiol.* 82, 205–233. doi: 10.1016/S0074-7742(07)82011-3

Butterfield, D., and Halliwell, B. (2019). Oxidative stress, dysfunctional glucose metabolism and Alzheimer disease. *Nat. Rev. Neurosci.* 20, 148–160. doi: 10.1038/s41583-019-0132-6

Calsolaro, V., and Edison, P. (2016). Neuroinflammation in Alzheimer's disease: current evidence and future directions. *Alzheimer's Dement.* 12, 719–732. doi: 10.1016/j.jalz.2016.02.010

Caroni, P., and Grandes, P. (1990). Nerve sprouting in innervated adult skeletal muscle induced by exposure to elevated levels of insulin-like growth factors. *J. Cell Biol.* 110, 1307–1317. doi: 10.1083/jcb.110.4.1307

Carvalho, C., Cardoso, S., Correia, S., and Moreira, P. (2019). Tortuous paths of insulin signaling and mitochondria in Alzheimer's disease. *Adv. Exp. Med. Biol.* 1128, 161–183. doi: 10.1007/978-981-13-3540-2_9

De Felice, F., Gonçalves, R., and Ferreira, S. (2022). Impaired insulin signalling and allostatic load in Alzheimer disease. *Nat. Rev. Neurosci.* 23, 215–230. doi: 10.1038/s41583-022-00558-9

de la Monte, S. (2017). Insulin resistance and neurodegeneration: progress towards the development of new therapeutics for Alzheimer's disease. *Drugs* 77, 47–65. doi: 10.1007/s40265-016-0674-0

DeKosky, S., and Scheff, S. (1990). Synapse loss in frontal cortex biopsies in Alzheimer's disease: correlation with cognitive severity. *Ann. Neurol.* 27, 457–464. doi: 10.1002/ana.410270502

Eraky, S., and Ramadan, N. (2022). Abo El-Magd NF. Ameliorative effects of bromelain on aluminum-induced Alzheimer's disease in rats through modulation of TXNIP pathway. *Int. J. Biol. Macromol.* 227, 1119–1131. doi: 10.1016/j.ijbiomac.2022.11.291

Fernandez, A., and Torres-Alemán, I. (2012). The many faces of insulin-like peptide signalling in the brain. *Nat. Rev. Neurosci.* 13, 225–239. doi: 10.1038/nrn3209

Fuster, J. (1993). Frontal lobes. *Curr. Opin. Neurobiol.* 3, 160–165. doi: 10.1016/0959-4388(93)90204-C

Georges, J., Miller, O., and Bintener, C. (2020). *Estimating the prevalence of dementia in Europe.* Luxembourg: Alzheimer Europe

Hashimoto, A., Matsuoka, K., Yasuno, F., Takahashi, M., Iida, J., Jikumaru, K., et al. (2017). Frontal lobe function in elderly patients with Alzheimer's disease and caregiver burden. *Psychogeriatrics* 17, 267–272. doi: 10.1111/psyg.12231

He, L., Liu, X., Li, H., Dong, R., Liang, R., and Wang, R. (2022). Polyrhachis vicina roger alleviates memory impairment in a rat model of Alzheimer's disease through the EGR1/BACE1/APP axis. *ACS Chem. Neurosci.* 13, 1857–1867. doi: 10.1021/acschemneuro.1c00193

Hong, S., Beja-Glasser, V., Nfonoyim, B., Frouin, A., Li, S., Ramakrishnan, S., et al. (2016). Complement and microglia mediate early synapse loss in Alzheimer mouse models. *Science* 352, 712–716. doi: 10.1126/science.aad8373

Hu, Y., Chen, X., Huang, S., Zhu, Q., Yu, S., Shen, Y., et al. (2019). Early growth response-1 regulates acetylcholinesterase and its relation with the course of Alzheimer's disease. *Brain Pathol.* 29, 502–512. doi: 10.1111/bpa.12688

Jiang, T., Sun, Q., and Chen, S. (2016). Oxidative stress: a major pathogenesis and potential therapeutic target of antioxidative agents in Parkinson's disease and Alzheimer's disease. *Prog. Neurobiol.* 147, 1–19. doi: 10.1016/j.pneurobio.2016.07.005

Johnson, W., Li, C., and Rabinovic, A. (2007). Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* 8, 118–127. doi: 10.1093/biostatistics/kxj037

Journal of Nature Genetics (2019). Deep learning for genomics. *Nat. Genet.* 51:1. doi: 10.1038/s41588-018-0328-0

Kao, Y., Ho, P., Tu, Y., Jou, I., and Tsai, K. (2020). Lipids and Alzheimer's disease. *Int. J. Mol. Sci.* 21:1505. doi: 10.3390/ijms21041505

Katzman, R. (2008). The prevalence and malignancy of Alzheimer disease: a major killer. *Alzheimers Dement.* 4, 378–380. doi: 10.1016/j.jalz.2008.10.003

Kellar, D., and Craft, S. (2020). Brain insulin resistance in Alzheimer's disease and related disorders: mechanisms and therapeutic approaches. *Lancet Neurol.* 19, 758–766. doi: 10.1016/S1474-4422(20)30231-3

Kim, N., Yu, L., Dawe, R., Petyuk, V., Gaiteri, C., De Jager, P., et al. (2019). Microstructural changes in the brain mediate the association of AK4, IGFBP5, HSPB2, and ITPK1 with cognitive decline. *Neurobiol. Aging.* 84, 17–25. doi: 10.1016/j.neurobiolaging.2019.07.013

Koldamova, R., Schug, J., Lefterova, M., Cronican, A., Fitz, N., Davenport, F., et al. (2014). Genome-wide approaches reveal EGR1-controlled regulatory networks associated with neurodegeneration. *Neurobiol. Dis.* 63, 107–114. doi: 10.1016/j.nbd.2013.11.005

Kosenko, E., Tikhonova, L., Alilova, G., Urios, A., and Montoliu, C. (2020). The erythrocytic hypothesis of brain energy crisis in sporadic alzheimer disease: possible consequences and supporting evidence. *J. Clin. Med.* 9:206. doi: 10.3390/jcm9010206

Langfelder, P., Horvath, S., Langfelder, P., and Horvath, S. (2009). WGCNA: an R package for weighted correlation network analysis. *BMC Bioinform.* 9:559. doi: 10.1186/1471-2105-9-559

Li, R., Li, L., Xu, Y., and Yang, J. (2022). Machine learning meets omics: applications and perspectives. *Brief. Bioinform.* 23:bbab460. doi: 10.1093/bib/bbab460

Lim, A., Gaiteri, C., Yu, L., Sohail, S., Swardfager, W., Tasaki, S., et al. (2018). Seasonal plasticity of cognition and related biological measures in adults with and without Alzheimer disease: analysis of multiple cohorts. *PLoS Med.* 15:e1002647. doi: 10.1371/journal.pmed.1002647

Lustbader, W. J. (2004). ABAD directly links a? To mitochondrial toxicity in Alzheimer\"s disease. *Science* 304, 448–452. doi: 10.1126/science.1091230

Marshall, G., Gatchel, J., Donovan, N., Muniz, M., Schultz, A., Becker, J., et al. (2019). Regional tau correlates of instrumental activities of daily living and apathy in mild cognitive impairment and Alzheimer's disease dementia. *J. Alzheimers Dis.* 67, 757–768. doi: 10.3233/JAD-170578

Mckhann, G., Knopman, D., Chertkow, H., Hyman, B., Jack, C., Kawas, C., et al. (2011). The diagnosis of dementia due to Alzheimer's disease: recommendations from the national institute on aging-Alzheimer's association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's Dement.* 7, 263–269. doi: 10.1016/j.jalz.2011.03.005

Minoru, K., Miho, F., Mao, T., Yoko, S., and Kanae, M. (2017). KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* 45, D353–D361. doi: 10.1093/nar/gkw1092

Mukherjee, S., Russell, J., Carr, D., Burgess, J., Allen, M., Serie, D., et al. (2017). Systems biology approach to late-onset Alzheimer's disease genome-wide association study identifies novel candidate genes validated using brain expression data and Caenorhabditis elegans experiments. *Alzheimers Dement.* 13, 1133–1142. doi: 10.1016/j.jalz.2017.01.016

Mulder, N., Adebiyi, E., Adebiyi, M., Adeyemi, S., Ahmed, A., Ahmed, R., et al. (2017). Development of bioinformatics infrastructure for genomics research. *Glob. Heart.* 12, 91–98. doi: 10.1016/j.gheart.2017.01.005

Nalbantoglu, J., Lacoste-Royal, G., and Gauvreau, D. (2005). Genetic factors in Alzheimer\"s disease. *J. Am. Geriatr. Soc.* 38, 564–568. doi: 10.1111/j.1532-5415.1990.tb02408.x

Perry, G., Moreira, P., Santos, M., Oliveira, C., Shenk, J., Nunomura, A., et al. (2008). Alzheimer disease and the role of free radicals in the pathogenesis of the disease. *Cns Neurol. Disord. Drug Targets* 7, 3–10. doi: 10.2174/187152708783885156

Pevsner, J. (2015). *Bioinformatics and functional genomics.* Hoboken, NJ: John Wiley & Sons.

Querfurth, H., and Laferla, F. (2010). Alzheimer's disease. *N. Engl. J. Med.* 362, 329–344. doi: 10.1056/NEJMra0909142

Rad, S., Arya, A., Karimian, H., Madhavan, P., Rizwan, F., Koshy, S., et al. (2018). Mechanism involved in insulin resistance via accumulation of beta-amyloid and neurofibrillary tangles: link between type 2 diabetes and Alzheimer's disease. *Drug Des. Deve. Ther.* 12, 3999–4021. doi: 10.2147/DDDT.S173970

Rajendran, L., and Paolicelli, R. (2018). Microglia-mediated synapse loss in Alzheimer's disease. *J. Neurosci.* 38, 2911–2919. doi: 10.1523/JNEUROSCI.1136-17.2017

Ranstam, J., and Cook, J. (2018). LASSO regression. *Br. J. Surg.* 105:1348. doi: 10.1002/bjs.10895

Rauskolb, S., Andreska, T., Fries, S., von Collenberg, C., Blum, R., Monoranu, C., et al. (2022). Insulin-like growth factor 5 associates with human Aß plaques and promotes cognitive impairment. *Acta Neuropathol. Commun.* 10:68. doi: 10.1186/s40478-022-01352-5

Rauskolb, S., Dombert, B., and Sendtner, M. (2017). Insulin-like growth factor 1 in diabetic neuropathy and amyotrophic lateral sclerosis. *Neurobiol. Dis.* 97(Pt B), 103–113. doi: 10.1016/j.nbd.2016.04.007

Rigatti, S. (2017). Random forest. *J. Insur. Med.* 47, 31–39. doi: 10.17849/insm-47-01-31-39.1

Robinson, M., Lee, B., and Hane, F. (2017). Recent progress in Alzheimer's disease research. Part 2: genetics and epidemiology. *J. Alzheimer's Dis.* 57, 317–330. doi: 10.3233/JAD-161149

Salehi, Z., Mashayekhi, F., and Naji, M. (2008). Insulin like growth factor-1 and insulin like growth factor binding proteins in the cerebrospinal fluid and serum from patients with Alzheimer's disease. *Biofactors* 33, 99–106. doi: 10.1002/biof.5520330202

Sanz, H., Valim, C., Vegas, E., Oller, J., and Reverter, F. (2018). SVM-RFE: selection and visualization of the most relevant features through non-linear kernels. *BMC Bioinform.* 19:432. doi: 10.1186/s12859-018-2451-4

Sbai, O., Djelloul, M., Auletta, A., Ieraci, A., Vascotto, C., and Perrone, L. (2022). AGE-TXNIP axis drives inflammation in Alzheimer's by targeting Aβ to mitochondria in microglia. *Cell Death Dis.* 13:302. doi: 10.1038/s41419-022-04758-0

Seshan, V., Gnen, M., and Begg, C. (2013). Comparing ROC curves derived from regression models. *Stat. Med.* 32, 1483–1493. doi: 10.1002/sim.5648

Smith, M. (1998). Alzheimer disease. *Int. Rev. Neurobiol.* 42, 1–54. doi: 10.1016/S0074-7742(08)60607-8

Sosa-Ortiz, A., Acosta-Castillo, I., and Prince, M. (2012). Epidemiology of dementias and Alzheimer's disease. *Arch. Med. Res.* 43, 600–608. doi: 10.1016/j.arcmed.2012.11.003

Stoeckel, L., Stewart, C., Griffith, H., Triebel, K., Okonkwo, O., den Hollander, J., et al. (2013). MRI volume of the medial frontal cortex predicts financial capacity in patients with Alzheimer's disease. *Brain Imaging Behav.* 7, 282–292. doi: 10.1007/s11682-013-9226-3

Subramanian, A., Tamayo, P., Mootha, V., Mukherjee, S., Ebert, B., Gillette, M., et al. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U.S.A.* 102, 15545–15550. doi: 10.1073/pnas.0506580102

Swerdlow, R., Burns, J., and Khan, S. (2014). The Alzheimer's disease mitochondrial cascade hypothesis: progress and perspectives. *Biochim. Biophys. Acta Mol. Basis Dis.* 1842, 1219–1231. doi: 10.1016/j.bbadis.2013.09.010

Tabaton, M., Cammarata, S., Mancardi, G., Manetto, V., Autilio-Gambetti, L., Perry, G., et al. (1991). Ultrastructural localization of beta-amyloid, tau, and ubiquitin epitopes in extracellular neurofibrillary tangles. *Proc. Natl. Acad. Sci. U.S.A.* 88, 2098–2102. doi: 10.1073/pnas.88.6.2098

Tran, K., Kondrashova, O., Bradley, A., Williams, E., Pearson, J., and Waddell, N. (2021). Deep learning in cancer diagnosis, prognosis and treatment selection. *Genome Med.* 13:152. doi: 10.1186/s13073-021-00968-x

Tsubaki, H., Tooyama, I., and Walker, D. (2020). Thioredoxin-interacting protein (TXNIP) with focus on brain and neurodegenerative diseases. *Int. J. Mol. Sci.* 21:9357. doi: 10.3390/ijms21249357

Tublin, J., Adelstein, J., Del Monte, F., Combs, C., and Wold, L. (2019). Getting to the heart of Alzheimer disease. *Circ. Res.* 124, 142–149. doi: 10.1161/CIRCRESAHA.118.313563

Wang, C., Xu, Y., Wang, X., Guo, C., Wang, T., and Wang, Z. (2019). Dl-3-n-butylphthalide inhibits NLRP3 inflammasome and mitigates Alzheimer's-like pathology via Nrf2-TXNIP-TrX axis. *Antioxid Redox Signal.* 30, 1411–1431. doi: 10.1089/ars.2017.7440

Wyss-Coray, T. (2006). Inflammation in Alzheimer disease: driving force, bystander or beneficial response? *Nat. Med.* 12, 1005–1015.

Xiao, Z., Chu, Y., and Qin, W. (2020). IGFBP5 modulates lipid metabolism and insulin sensitivity through activating AMPK pathway in non-alcoholic fatty liver disease. *Life Sci.* 256:117997. doi: 10.1016/j.lfs.2020.117997

Yu, G., and He, Q. (2016). ReactomePA: an R/Bioconductor package for reactome pathway analysis and visualization. *Mol. Biosyst.* 12, 477–479. doi: 10.1039/C5MB00663E

Yu, G., Wang, L., Han, Y., and He, Q. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics* 16, 284–287. doi: 10.1089/omi.2011.0118

Yu, L., Petyuk, V., Gaiteri, C., Mostafavi, S., Young-Pearse, T., Shah, R., et al. (2018). Targeted brain proteomics uncover multiple pathways to Alzheimer's dementia. *Ann. Neurol.* 84, 78–88. doi: 10.1002/ana.25266

Zhang, M., Hu, G., Shao, N., Qin, Y., Chen, Q., Wang, Y., et al. (2021). Thioredoxin-interacting protein (TXNIP) as a target for Alzheimer's disease: flavonoids and phenols. *Inflammopharmacology* 29, 1317–1329. doi: 10.1007/s10787-021-00861-4

Zhao, Y., and Zhao, B. (2013). Oxidative stress and the pathogenesis of Alzheimer's disease. *Oxid. Med. Cell Longev.* 2013:316523.

Zhu, Q., Unmehopa, U., Bossers, K., Hu, Y., Verwer, R., Balesar, R., et al. (2016). MicroRNA-132 and early growth response-1 in nucleus basalis of Meynert during the course of Alzheimer's disease. *Brain* 139(Pt 3), 908–921. doi: 10.1093/brain/awv383