frontiers | Frontiers in Aging Neuroscience

TYPE Original Research
PUBLISHED 16 June 2023
DOI 10.3389/fnagi.2023.1161068

OPEN ACCESS

EDITED BY
Fereshteh Farajdokht,
Tabriz University of Medical Sciences, Iran

REVIEWED BY
Chunlei Zheng,
Boston University, United States
Deepika Mathur,
Icahn School of Medicine at Mount Sinai, United States
Naazneen Khan,
University of Kentucky, United States
Haoxiang Cheng,
Icahn School of Medicine at Mount Sinai, United States

*CORRESPONDENCE
Xueqin Sui
✉ suixueqin2023@126.com

†These authors have contributed equally to this work

RECEIVED 07 February 2023
ACCEPTED 30 May 2023
PUBLISHED 16 June 2023

CITATION
Ma C, Li J, Chi Y, Sun X, Yang M and Sui X (2023) Identification and prediction of m7G-related Alzheimer's disease subtypes: insights from immune infiltration and machine learning models.
Front. Aging Neurosci. 15:1161068.
doi: 10.3389/fnagi.2023.1161068

Identification and prediction of m7G-related Alzheimer's disease subtypes: insights from immune infiltration and machine learning models

Chao Ma[1,2†], Jian Li[3†], Yuhua Chi[1], Xuan Sun[1], Maoquan Yang[2] and Xueqin Sui[1*]

[1]Department of General Medicine, Affiliated Hospital of Weifang Medical University, Weifang, Shandong, China, [2]School of Clinical Medicine, Weifang Medical University, Weifang, Shandong, China, [3]Department of Neurology, Affiliated Hospital of Weifang Medical University, Weifang, Shandong, China

Introduction: Alzheimer's disease (AD) is a complex and progressive neurodegenerative disorder that primarily affects older individuals. N7-methylguanosine (m7G) is a common RNA chemical modification that impacts the development of numerous diseases. Thus, our work investigated m7G-related AD subtypes and established a predictive model.

Methods: The datasets for AD patients, including GSE33000 and GSE44770, were obtained from the Gene Expression Omnibus (GEO) database, which were derived from the prefrontal cortex of the brain. We performed differential analysis of m7G regulators and examined the immune signatures differences between AD and matched-normal samples. Consensus clustering was employed to identify AD subtypes based on m7G-related differentially expressed genes (DEGs), and immune signatures were explored among different clusters. Furthermore, we developed four machine learning models based on the expression profiles of m7G-related DEGs and identified five important genes from the optimal model. We evaluated the predictive power of the 5-gene-based model using an external AD dataset (GSE44770).

Results: A total of 15 genes related to m7G were found to be dysregulated in patients with AD compared to non-AD patients. This finding suggests that there are differences in immune characteristics between these two groups. Based on the differentially expressed m7G regulators, we categorized AD patients into two clusters and calculated the ESTIMATE score for each cluster. Cluster 2 exhibited a higher ImmuneScore than Cluster 1. We performed the receiver operating characteristic (ROC) analysis to compare the performance of four models, and we found that the Random Forest (RF) model had the highest AUC value of 1.000. Furthermore, we tested the predictive efficacy of a 5-gene-based RF model on an external AD dataset and obtained an AUC value of 0.968. The nomogram, calibration curve, and decision curve analysis (DCA) confirmed the accuracy of our model in predicting AD subtypes.

Conclusion: The present study systematically examines the biological significance of m7G methylation modification in AD and investigates its association with immune infiltration characteristics. Furthermore, the study

develops potential predictive models to assess the risk of m7G subtypes and the pathological outcomes of patients with AD, which can facilitate risk classification and clinical management of AD patients.

KEYWORDS

Alzheimer's disease, m7G methylation, AD subtypes, immune infiltration, machine learning, predictive model

Introduction

Alzheimer's disease (AD) is a neurodegenerative disorder that affects the central nervous system and results in a progressive decline in cognitive function and memory. It predominantly affects the elderly population and significantly impacts patients' quality of life (Cao and Zheng, 2018; Liu et al., 2020; Qiu et al., 2022). Recent statistics indicate that approximately 50 million individuals worldwide suffer from AD and other forms of dementia, and the number is increasing each year, resulting in a substantial burden on patients, families, society, and the healthcare system (Suresh et al., 2021). However, the pathogenesis of AD is exceedingly intricate, and its exact causative factors remain to be fully elucidated, impeding the development of AD drugs (Mangialasche et al., 2010; Ballard et al., 2011). Consequently, current treatments for AD are unsatisfactory. Nevertheless, recent advancements in bioinformatics, particularly leveraging the GEO database, have enabled the in-depth exploration of biomarkers that contribute to AD development, facilitating the development of multi-factor prediction models (Zhang T. et al., 2021; Li et al., 2022; Lin et al., 2022; Wang et al., 2022). These models can provide new insights into the individualized and precise treatment of AD patients.

There is mounting evidence to suggest that epigenetic modifications participate in the development of diseases by regulating gene expression post-transcriptionally (Zhang M. et al., 2021; Cheng et al., 2022; Luo et al., 2022). N7-methylguanosine (m7G) is a common RNA methylation that plays a crucial role in maintaining RNA processing, metabolism, stability, nuclear export, and translation (Wei et al., 2022; Zhong et al., 2022). Recent studies have shown a strong association between m7G and various pathological processes that affect the diagnosis and prognosis of diseases (Chen et al., 2022; Wei et al., 2022; Xia et al., 2023).

The role of the brain's immune system, specifically the microglia, has been extensively investigated in the development of Alzheimer's disease. Microglia, a specialized type of immune cell, play a crucial role in clearing debris and toxic substances from the brain (Jevtic et al., 2017). Identifying the precise role of microglia will aid researchers in developing more effective strategies to target this system, which could potentially prevent the disease from progressing in the early stages.

Several studies have indicated that m7G methylation is involved in the immune response to several diseases. RNA methylation in the immune system influences the maturation and response functions of immune cells (Lu et al., 2022). m7G modifications regulate innate immunity by influencing RNA immunogenicity and innate immune components in the body. Methylation modification patterns mediated by m7G regulators may be associated with tumor microenvironment infiltration in glioblastoma (Wu et al., 2022).

This study aimed to conduct a differential analysis of m7G regulators in AD and non-AD samples and examine the differences in immune features. Based on 15 differentially expressed m7G-related genes (M7RGs), we identified two m7G-related subtypes and assessed the differences in immune features between the subtypes. To investigate the biological processes involving m7G-related DEGs between the subtypes, we performed Gene Set Variation Analysis (GSVA). Additionally, we constructed four machine-learning models using m7G-related DEGs and evaluated the diagnostic power of multiple models using ROC curves.

Materials and methods

Data collection

Two datasets related to AD were collected from the GEO website database,[1] namely GSE33000 and GSE44770. Both datasets contained samples obtained from human prefrontal cortex brain tissues. The GSE33000 dataset (GPL4372 platform) consisted of 157 healthy and 310 AD samples, which served as the training group. On the other hand, the GSE44770 dataset (GPL4372 platform) comprised 129 AD and 101 normal samples, and it was used as the test group. The gene expression profiles of the three datasets were normalized and processed using the "Perl" script and R package "limma." Subsequently, 22 m7G regulators were identified from previous publications (Luo et al., 2022; Li et al., 2023; Maimaiti et al., 2023; Qin et al., 2023).

Immune infiltration analysis

The study utilized the CIBERSORT algorithm to evaluate the relative abundances and infiltration scores of 22 immune cell types in AD samples, based on gene expression profiles (Gentles et al., 2015). To explore the relationship between M7RGs and immune cell types, correlation analysis was employed. A significance level of

Abbreviations: AD, Alzheimer's disease; PCA, component analysis; CM, consensus matrix; CDF, cumulative distribution function; DCA, decision curve analysis; DEGs, differentially expressed genes; DMRs, differentially methylated regions; XGB, eXtreme gradient boosting; GEO, gene expression omnibus; GLM, generalized linear model; GSVA, gene set variation analysis; LIHC, hepatocellular carcinoma; m7G, N7-methylguanosine; PAAD, pancreatic cancer; PAH, pulmonary arterial hypertension; RF, random forest; ROC, receiver operating characteristic; Tregs, regulatory T cells; RMSE, root mean square error; SVM, support vector machines.

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

$P < 0.05$ was used. The results were presented using the R packages "reshape2" and "ggpubr."

Consensus clustering for AD patients

Based on the landscape of DEGs expression associated with AD, we classified AD samples into distinct m7G-related subtypes by utilizing the R package "Consensus Cluster Plus." The maximum cluster number, $k = 9$ was selected, and the optimal cluster number was evaluated based on the consensus matrix (CM) and CDF. To evaluate the distribution between m7G-related clusters, we utilized Principal Component Analysis (PCA).

GSVA

In this study, we employed GSVA, a differential analysis approach at the pathway level, to investigate the discrepancies in biological activities among the clusters of M7RGs. We implemented the GSVA method using the R package "GSVA." We obtained GSVA gene sets from the "curated gene sets" and "ontology gene sets" modules available in the MSigDB database.

Construction of machine-learning models and a nomogram

The cluster-specific DEGs were identified by intersecting the AD-related hub genes and the m7G cluster-related hub genes. Subsequently, the expression patterns of these DEGs were analyzed to construct four machine learning models using the R package "caret." The models included the Random Forest (RF) (Rigatti, 2017), SVM (Gold and Sollich, 2003), Generalized Linear Model (GLM) (Nelder and Wedderburn, 1972), and eXtreme Gradient Boosting (XGB) (Romeo and Frontoni, 2022).

The study considered the distinct clusters as the response variable and selected the cluster-specific DEGs as explanatory variables. A total of 310 AD samples were randomly divided into a training set ($N = 218$, 70%) and a validation set ($N = 92$, 30%). The R package "caret" was used to control the training process with the parameter set to fivefold repeated cross-validation. Then, the "randomForest," "kernlab," and "xgboost" packages were loaded in sequence, and four machine learning algorithms, namely, RF, SVM, GLM, and XGB, were chosen. The "explain" function in the "DALEX" package was used to evaluate the performance of these models and generate various indicators related to the prediction results. The "DALEX" package was also utilized to visualize the residual distribution and feature importance among these models, while the "pROC" package was used to plot the area under the ROC curves. The top five predictive genes associated with AD were identified based on the optimal machine learning model. Finally, the diagnostic value of the model was verified using the ROC curve analysis in the GSE44770 dataset.

We have developed a nomogram to predict the risk of AD patients by utilizing the five essential genes identified from the RF model. To assess the predictive performance of the nomogram, we utilized decision curve analysis and a calibration curve. These methods were employed to validate the efficacy of the predictive model.

Independent validation analysis

We opted to use the AD dataset GSE44770 to assess the diagnostic efficiency of a 5-gene-based RF model. We presented the outcomes in the form of ROC curves. Additionally, we developed a nomogram employing the GSE44770 dataset to assess the risk of m7G subtypes. The calibration curve and DCA were employed to evaluate the predictive performance of the nomogram model.

Statistical analysis

The statistical analysis was performed using R software (Version 4.2.1), with the Perl and "limma" packages used for data processing. The sample classification was carried out using the "Consensus Cluster Plus" package. For continuous variables, normality was assessed, and either the Student's *t*-test or Wilcoxon rank-sum test was employed for analysis. The Chi-square test was used for categorical variable differences. A two-sided adjusted *p*-value of < 0.05 was considered statistically significant.

Results

Identification of differentially expressed M7RGs

A detailed flow chart of the study process is exhibited in **Figure 1**. In order to investigate the biological significance of m7G in AD, we examined the expression profiles of 22 M7RGs and performed differential analysis on AD and normal samples using the GSE33000 dataset. From this analysis, we identified 15 differentially expressed M7RGs, with 4 of them (NUDT3, CYFIP1, NCBP1, and IFIT5) being upregulated in AD patients, while 11 of them (METTL1, DCPS, NUDT10, NUDT11, EIF4E, EIF4E2, EIF4E3, LARP1, EIF4G3, LSM1, and NCBP2L) were downregulated compared to non-AD patients (**Figures 2A, B, E**). In addition, we conducted a correlation analysis to explore the relationship between these differentially expressed M7RGs. Our results indicated that EIF4E3 was significantly positively correlated with NUDT11 and negatively correlated with CYFIP1 (**Figures 2C, D**).

Immune landscape analysis

The CIBERSORT analysis was utilized to assess the differences in immune features between AD and normal samples (**Figure 3A**). The results of our study revealed that AD samples had elevated levels of infiltrating naive $CD4^+$ T cells, resting NK cells, monocytes, M0, M2 macrophages, and neutrophils, whereas infiltration levels of plasma cells, $CD8^+$ T cells, activated NK cells, and eosinophils were decreased (**Figure 3B**). These findings suggest

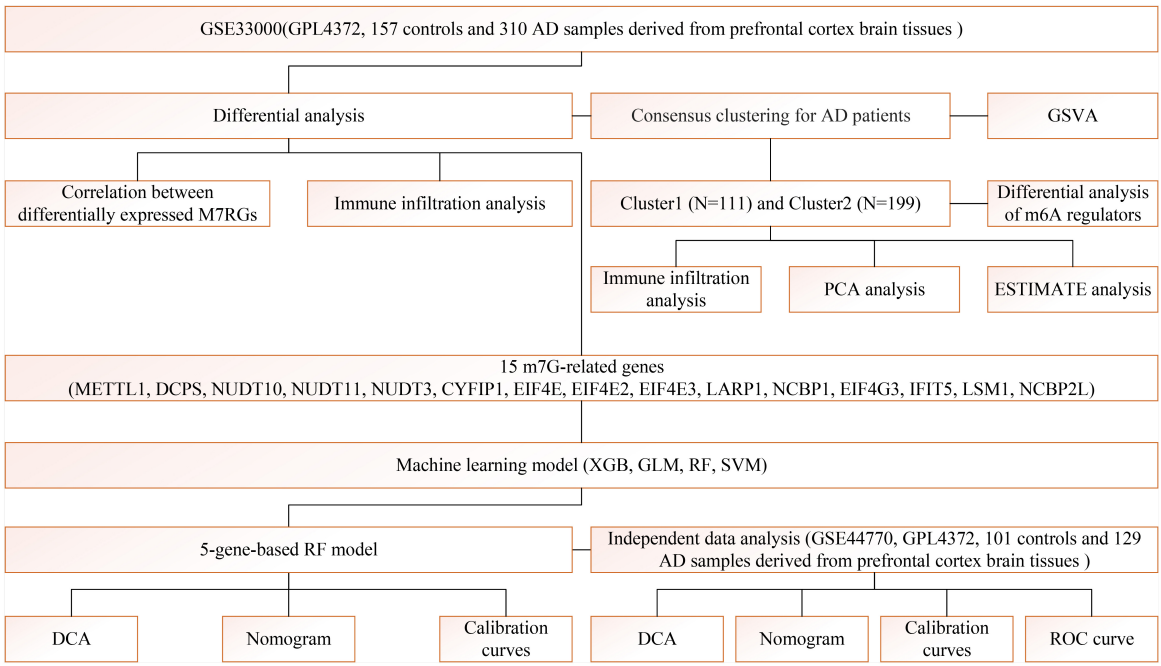

FIGURE 1
The study flow chart.

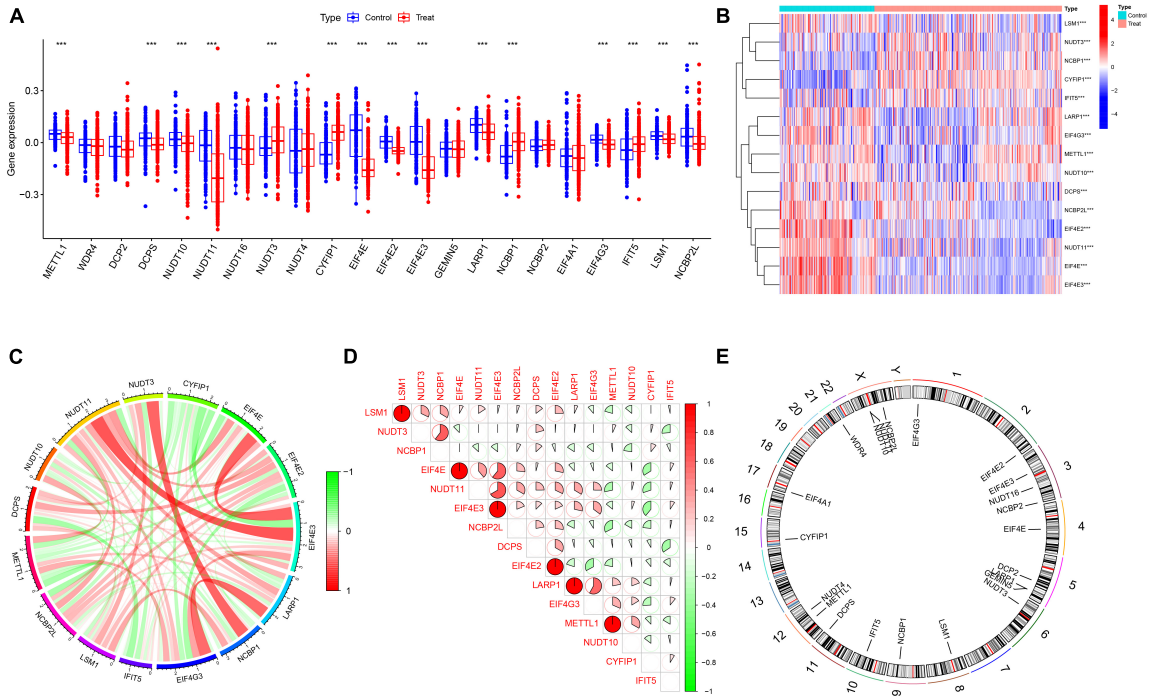

FIGURE 2
Identification of differentially expressed m7G-related genes. **(A)** Boxplots showed the expression of 22 M7RGs between AD and non-AD controls ***$p < 0.001$. **(B)** The expression patterns of 15 M7RGs were presented in the heatmap. **(C)** Interactions between 15 differentially expressed M7RGs at the molecular level. **(D)** Correlation analysis of 15 differentially expressed M7RGs. Red and green colors represent positive and negative correlations, respectively. The correlation coefficients were marked with the area of the pie chart. **(E)** The location of 22 M7RGs on chromosomes.

that alterations in the immune microenvironment play a crucial role in the development of AD. Moreover, the correlation analysis demonstrated that M2 macrophages exhibited a positive correlation with CYFIP1 and a negative correlation with METTL1. In contrast, neutrophils showed a positive association with the M7RGs (excluding LARP1), while M0 macrophages and naive $CD4^+$ T cells

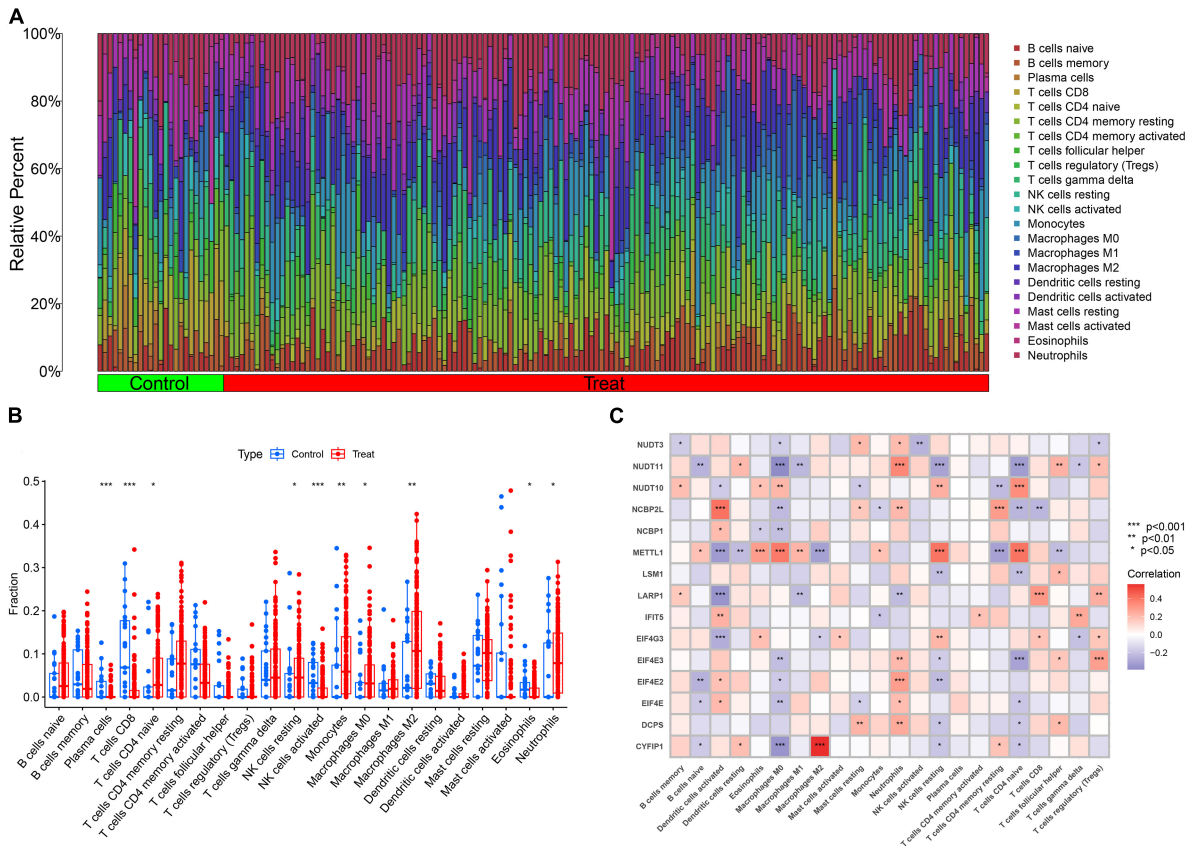

FIGURE 3
The analysis of immune features between AD and non-AD samples. **(A)** The relative abundances of 22 infiltrated immune cells between AD and non-AD controls. **(B)** Boxplots showed the differences in immune infiltration between AD and non-AD controls. **(C)** Correlation analysis between 15 M7RGs and 22 immune cell types. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

were negatively associated with the M7RGs (excluding METTL1 and NUDT1) (Figure 3C). These results suggest that M7RGs may participate in the onset and progression of AD by affecting immune cell infiltration levels. Moreover, the results of the CIBERSORTx analysis are shown in Supplementary Figure 1 and Supplementary Table 1.

Identification of m7G clusters in AD

Based on the differential expression of M7RGs, we conducted a consensus clustering analysis to investigate new molecular subtypes for patients with AD. The consensus clustering algorithm classified AD patients into two m7G-related subtypes when K was equal to 2 (Figures 4A–C). The PCA analysis revealed that the two clusters showed distinct molecular mechanisms related to m7G (Figure 4F).

Differential analysis of m7G regulators and immune signatures between m7G clusters

In this study, we performed a differential analysis of 15 M7RGs to investigate molecular signatures between two subtypes, Cluster 1 and Cluster 2. Our results showed that Cluster 1 exhibited elevated expression levels of DCPS, NUDT11, EIF4E, EIF4E2, EIF4E3, LARP1, EIF4G3, and NCBP2L, while Cluster 2 showed elevated expression levels of METTL1, CYFIP1, and NCBP1, as depicted in Figures 4D, E. Moreover, based on the CIBERSORT results, we observed a significant difference in immune cell infiltration between the two m7G-related subtypes, with Cluster 1 showing higher levels of regulatory T cells (Tregs) and neutrophils and Cluster 2 showing higher levels of naive $CD4^+$ T cells, resting NK cells, M0, and M1 macrophages (Figure 4H). We also conducted ESTIMATE analysis, which revealed that Cluster 2 had a higher StromalScore, ImmuneScore, and ESTIMATEScore, indicating elevated immune infiltration levels (Figures 4G, I).

Enrichment analysis

The study utilized GSVA analysis to investigate the differences in biological activities between two AD subtypes. The results showed that Cluster 1 had reinforced cell adhesion molecules, primary immunodeficiency, focal adhesion, and cancer-related pathways. In contrast, Cluster 2 showed activation of terpenoid backbone biosynthesis, regulation of autophagy, ubiquitin-mediated proteolysis, vibrio cholerae infection, and metabolism-related pathways (Figure 5A). Furthermore,

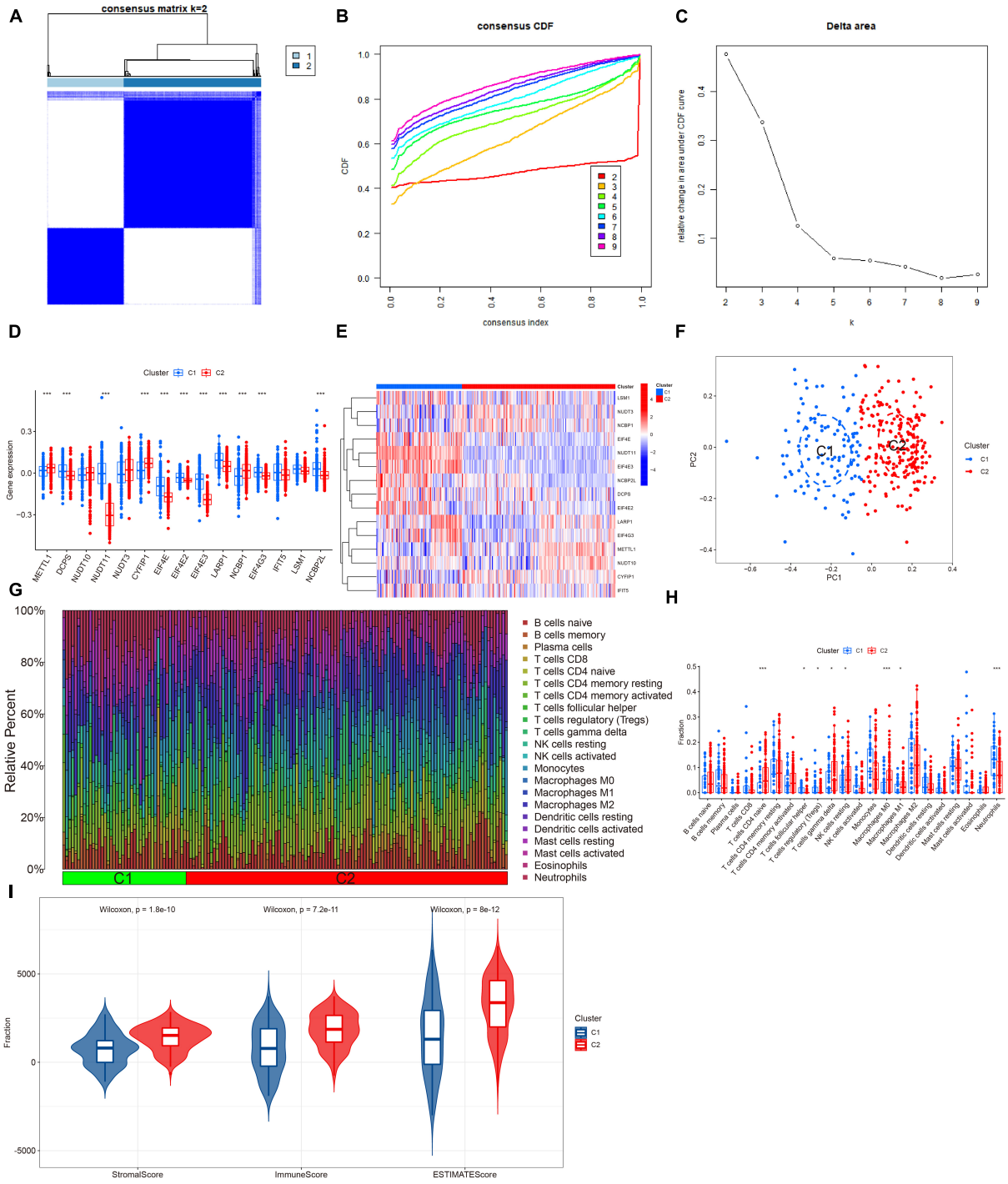

FIGURE 4
Identification of m7G-related molecular subtypes in AD. Consensus clustering matrix when $k = 2$ **(A)**, representative cumulative distribution function (CDF) curves **(B)**, CDF delta area curves **(C)**. **(D)** Boxplots showed the expression of 15 M7RGs between two m7G clusters $^{***}p < 0.001$. **(E)** The different analyses of 15 M7RGs between two m7G clusters were presented in the heatmap. **(F)** PCA analysis for cluster 1 and cluster 2. **(G)** The relative abundances of 22 infiltrated immune cells between two m7G clusters. **(H)** Boxplots showed the differences in immune infiltration between two m7G clusters $^{*}p < 0.05$, $^{***}p < 0.001$. **(I)** Violin plots showed the StromalScore, ImmuneScore, and ESTIMATEScore between two m7G clusters.

functional enrichment analysis revealed that Cluster 1 was significantly positively correlated with the secretion of lysosomal enzymes, the liposaccharide metabolic process, the translocon complex, the negative regulation of lipoprotein lipase activity, and lipid kinase activity. Conversely, Cluster 2 showed enrichment in golgi organization, nucleosome assembly, Wnt protein secretion, response to amphetamine, and regulation of cytoplasmic translation (Figure 5B). Thus, it was inferred that Cluster 2 may be involved in mRNA processing, translation, and metabolic pathways.

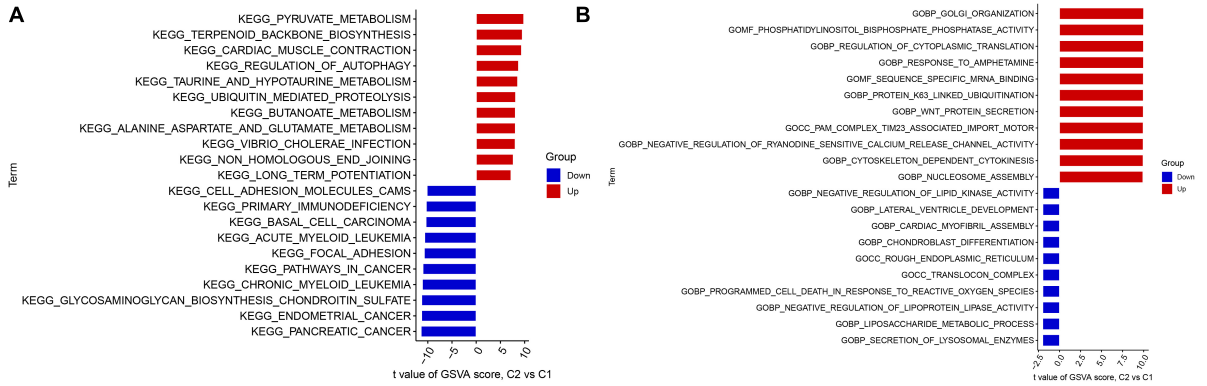

FIGURE 5
Identification of biological activities between two m7G clusters. **(A)** The KEGG pathway analysis. **(B)** The GO function analysis.

Construction of machine-learning models

To explore the m7G regulators associated with AD subtypes, we constructed four machine-learning models, including RF, SVM, GLM, and XGB using the 15 m7G-related DEGs in the AD training set. The XGB and GLM models showed lower residuals (**Figures 6A, B**). We identified 10 key genes from four modules, ranked by root mean square error (RMSE) (**Figure 6C**). In addition, we used the ROC curve to evaluate the diagnostic efficacy of the four models. The RF model exhibited the strongest diagnostic power (AUC = 1.000) (**Figure 6D**). Finally, we selected the top five best factors (NCBP2L, LARP1, EIF4E, EIF4E3, and NUDT11) from the RF model as predictor genes for further analysis.

We developed a nomogram to assess the predictive ability of the RF model across various AD subtypes (**Figure 7C**). To validate the predictive performance of the nomogram, we employed a calibration curve and DCA. The calibration curve showed that the predicted risk of AD clusters was nearly identical to the actual risk (**Figure 7A**). Furthermore, the DCA result demonstrated that the nomogram had a superior predictive ability, providing a theoretical basis for the prediction of m7G-related AD subtypes (**Figure 7B**). Subsequently, we classified AD patients into two m7G subtypes using a consensus clustering algorithm based on the GSE44770 dataset (**Figures 8A–C**). A nomogram was conducted to assess the risk of AD subtypes (**Figure 8F**). The calibration curve and DCA analysis were used to evaluate the diagnostic capability of the nomogram model (**Figures 8D, E**). The ROC analysis revealed an AUC of 0.968 for the 5 gene-based RF model in the GSE44770 dataset, demonstrating the higher diagnostic efficacy of our predictive model for AD subtypes (**Figure 8G**). Furthermore, the AUC values for NCBP2L, LARP1, EIF4E, EIF4E3, and NUDT11 were 0.700, 0.633, 0.861, 0.918, and 0.962, respectively (**Figure 8H**).

Discussion

The etiology of AD is complex, and its diagnosis is challenging. Consequently, AD patients often receive inadequate treatment. The traditional classification of AD based on histology poses a significant challenge to drug treatment due to drug resistance. Therefore, there is an urgent need to develop effective predictive models and molecular subtypes for risk stratification and personalized treatment of AD patients.

RNA methylation post-transcriptionally regulates target RNA metabolism and function, either promoting or inhibiting disease development (Dai et al., 2021; Cui et al., 2022). One of the most common chemical modifications is m7G, installed by key methyltransferases such as METTL1/WDR4, RAM/RNMT, and WBSCR22/TRMT112 (Luo et al., 2022). METTL1 has been identified as a high-risk gene in hepatocellular carcinoma (LIHC), contributing to tumor-associated phenotypes by inhibiting PTEN signaling (Tian et al., 2019). WDR4 plays a crucial role in promoting the proliferation of hepatocellular carcinoma by mediating m7G methylation (Xia et al., 2021). In pancreatic cancer (PAAD), the WBSCR22/TRMT112 complex downregulates the oncogene ISG15 expression, impairing malignant phenotypes (Khan et al., 2022).

The immune system is recognized as a crucial factor in AD, with various components of the immune system in both the brain and periphery interacting to contribute to the disease. The classical immune components of the nervous system, such as microglia and complement, as well as novel aspects of the peripheral immune system, such as monocytes and lymphocytes, are all implicated in AD.

Lu et al. (2022) found that most M7RGs are highly expressed in COVID-19 patients compared with that in non-COVID-19 patients. And m7G-cluster B showed higher immune infiltration and milder symptoms. Furthermore, m7G modification has been shown to play an essential role in the development of cardiovascular disease. Wei et al. (2022) indicated that m7G methylation is involved in the progression of pulmonary arterial hypertension (PAH) by affecting the immune microenvironment. However, the role of m7G in AD and its relationship with immunity remain incompletely understood.

Our objective was to investigate the biological activities of M7RGs in AD progression and identify m7G-related clusters and immune signatures. Additionally, we aimed to construct an optimal machine-learning model for evaluating the risk of AD subtypes. In this study, we obtained the expression profile of M7RGs from the GSE33000 dataset and performed differential analysis between

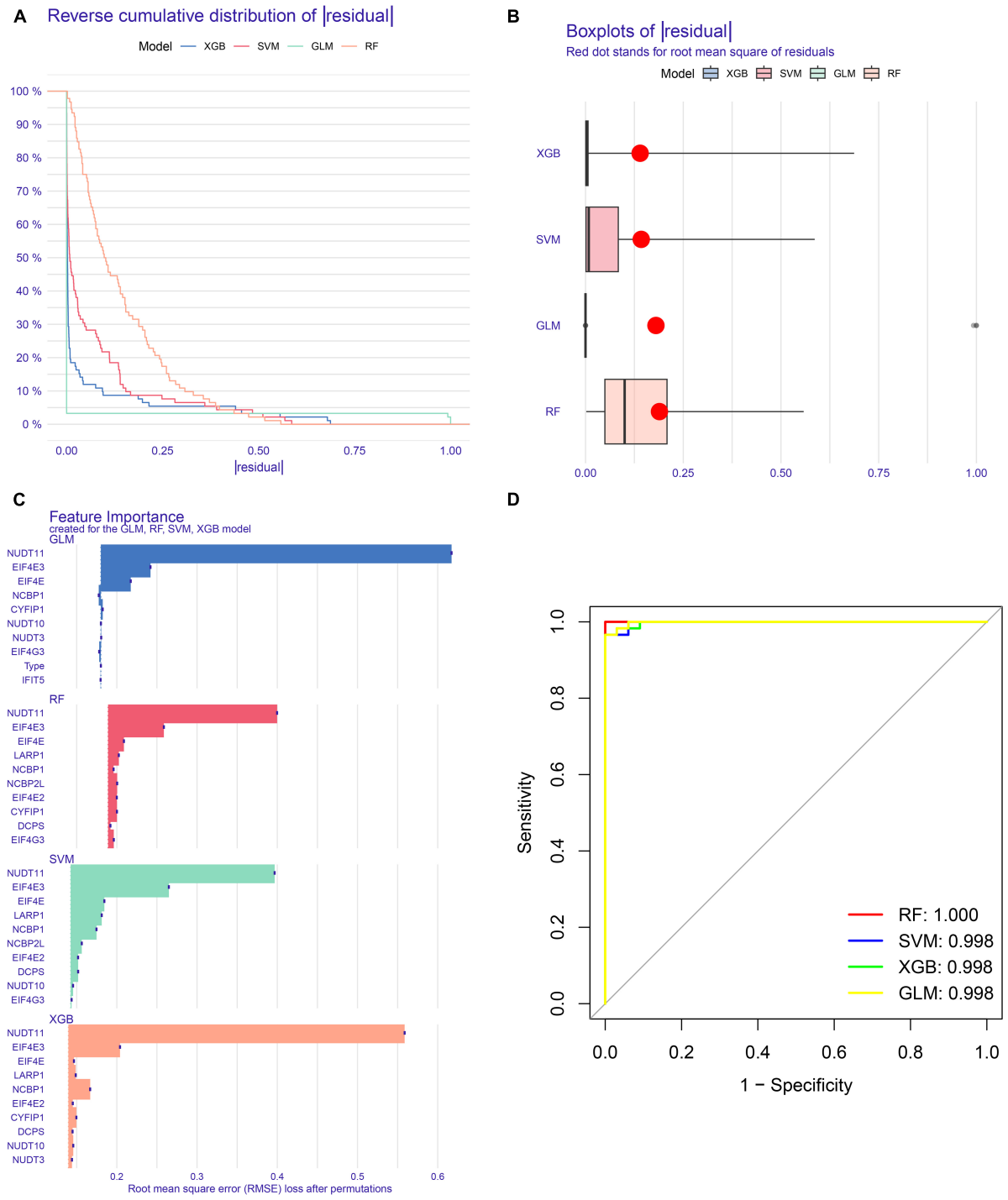

FIGURE 6
Construction and evaluation of RF, SVM, GLM, and XGB machine models. **(A)** Cumulative residual distribution of each machine learning model. **(B)** Boxplots showed the residuals of each machine learning model. Red dot represented the root mean square of residuals (RMSE). **(C)** The important features in RF, SVM, GLM, and XGB machine models. **(D)** ROC curves of four machine learning models based on fivefold repeated cross-validation in the testing cohort.

AD and non-AD individuals. Our data revealed dysregulation of 15 M7RGs in AD patients. A correlation analysis between m7G regulators was conducted to unveil the relationship between m7G regulators and AD. Furthermore, CIBERSORT analysis showed an immune infiltration difference between AD and non-AD patients. Infiltration levels of naive $CD4^+$ T cells, resting NK cells, neutrophils, and M0 and M2 macrophages were found to be elevated in AD patients.

The results from GSVA indicated that Cluster 1 was mainly enriched in cell adhesion molecules, primary immunodeficiency, focal adhesion, and cancer-related pathways. On the other hand, Cluster 2 was characterized by metabolism-related pathways,

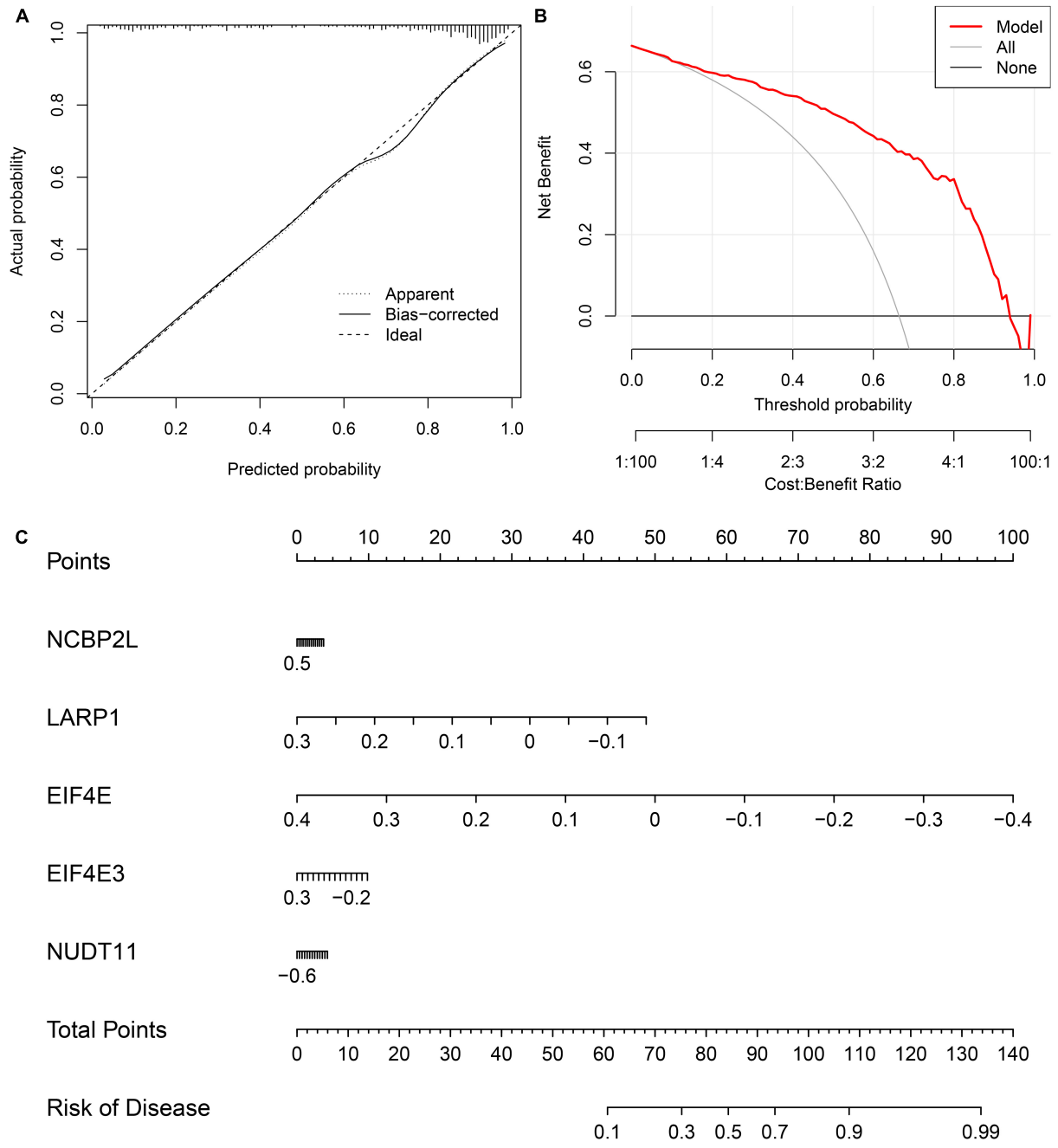

FIGURE 7
Validation of the 5-gene-based RF model. Construction of the calibration curve **(A)** and DCA **(B)** for assessing the predictive efficiency of the nomogram model. **(C)** Construction of a nomogram for predicting the risk of AD clusters based on the 5-gene-based RF model.

mRNA processing, and translation. Furthermore, Cluster 2 was enriched in terpenoid backbone biosynthesis, regulation of autophagy, ubiquitin-mediated proteolysis, vibrio cholerae infection, and metabolism-related pathways (Figure 7B). Additionally, functional enrichment analyses indicated that Cluster 1 was strongly associated with the secretion of lysosomal enzymes, the liposaccharide metabolic process, the translocon complex, the negative regulation of lipoprotein lipase activity, and lipid kinase activity. However, Cluster 2 was enriched in golgi organization, nucleosome assembly, Wnt protein secretion, the response to amphetamine, and the regulation of cytoplasmic translation (Figure 7C). Thus, we postulate that Cluster 2 may play a role in regulating amino acid metabolism-related pathways, mRNA processing, and translation. Notably, previous studies have reported that taurine and glutamate intake could improve learning and memory function and delay the progression of Alzheimer's disease. Consistent with these findings, Cluster 2 showed a stronger activity of taurine- and glutamate-related metabolism pathways. Taken together, these results suggest that Cluster 2 may have more activated amino acid metabolism-related pathways to halt the development of Alzheimer's disease.

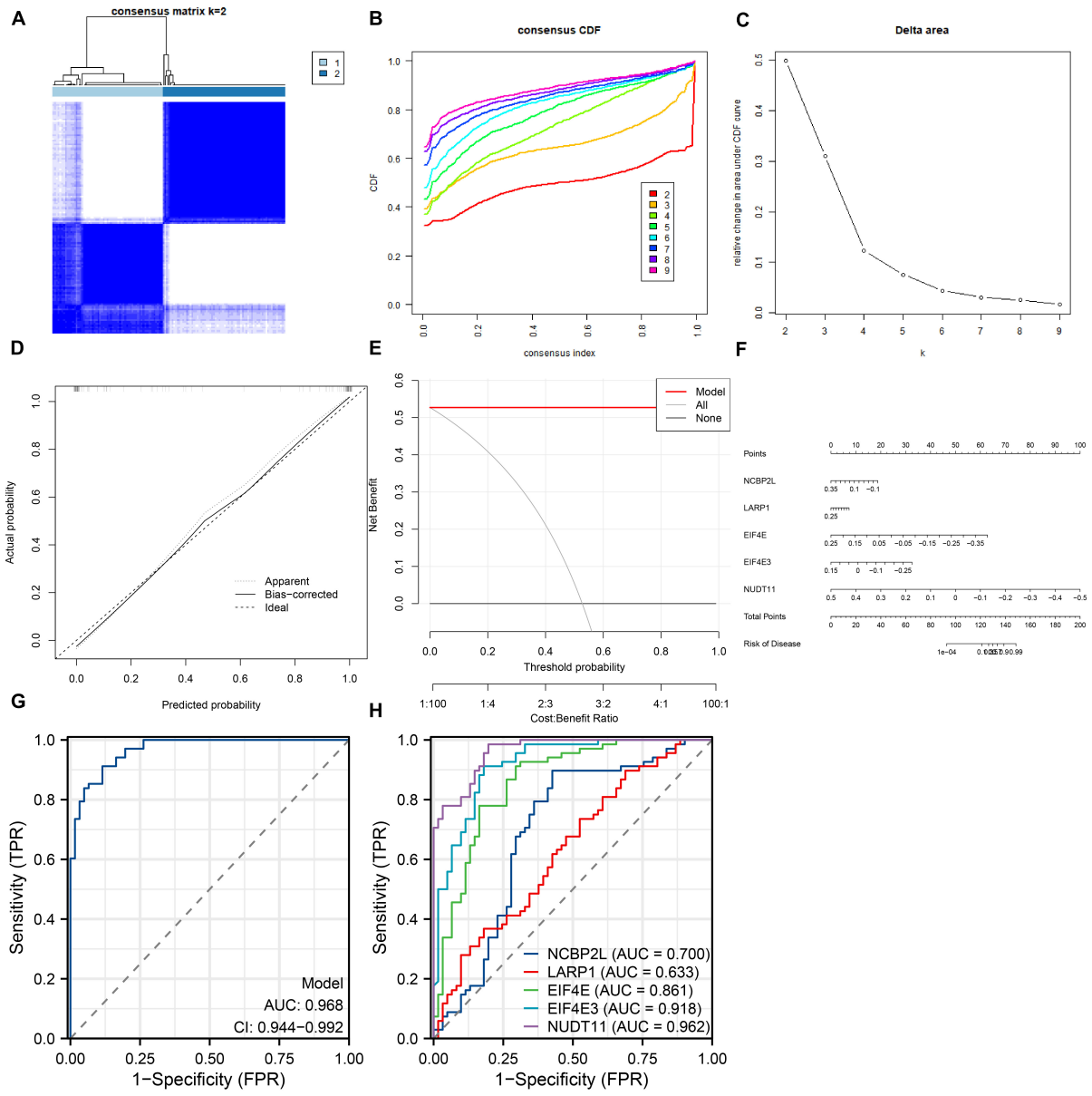

FIGURE 8
Independent validation analysis based on the GSE44770 dataset. Consensus clustering matrix when k = 2 **(A)**, representative cumulative distribution function (CDF) curves **(B)**, CDF delta area curves **(C)**. Construction of the calibration curve **(D)** and DCA **(E)** for assessing the predictive efficiency of the nomogram model. **(F)** Construction of a nomogram for predicting the risk of AD clusters based on the 5-gene-based RF model. **(G)** ROC curve of the 5-gene-based RF model. **(H)** ROC curves of NCBP2L, LARP1, EIF4E, EIF4E3, NUDT11.

Mining hub genes with higher diagnostic value for a disease using multiple machine learning models is becoming a common research method. In this study, we constructed four machine-learning models to screen five important genes associated with AD from 15 m7G-related DEGs. We used ROC curves to assess the predictive power of the four models, and the results suggested that the RF model exhibited relatively higher diagnostic performance (AUC = 1.000) and has the potential to distinguish between different AD subgroups. We established a nomogram to assess the risk of AD subtypes based on the expression profiles of five important genes, including NCBP2L, EIF4E, EIF4E3, LARP1, and NUDT11. The DCA and calibration curve further indicated that this model exhibited relatively higher diagnostic value and potential for clinical application.

Kenny et al. (2019) reported that EIF4E could serve as a specific biomarker for diagnosing AD. The phosphorylated levels of EIF4E were found to be significantly higher in patients with advanced AD, which correlated positively with the phosphorylated Tau protein levels. Therefore, the phosphorylated EIF4E protein might promote the progression of AD by enhancing the accumulation of phosphorylated Tau protein in neurons (Li et al., 2004). Furthermore, the study revealed that the level of differentially methylated regions (DMRs) was associated with various dementia-related genes, including EIF4E, that contribute to the onset of AD (Pérez et al., 2022).

To further validate the diagnostic efficacy of the 5-gene-based RF model, we selected an external AD dataset (GSE44770). The AUC value is 0.968 in GSE44770, suggesting that the model has potential for a wide range of applications in AD diagnosis.

Conclusion

Our study revealed that dysregulated m7G regulators are commonly found in patients with AD, which have an impact on the immune microenvironment. We identified two clusters of m7G and analyzed the differences in immune features between these two clusters. We chose a 5-gene-based RF model as the optimal machine learning model, which can accurately assess the risk of different subtypes of AD. Our findings elucidate the biological significance of m7G regulators in AD and provide a valuable insight for the risk stratification and clinical treatment of AD.

Data availability statement

The data presented in this study are deposited in the Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo/) repository, accession numbers: GSE33000 and GSE44770.

Author contributions

XS designed the study, reviewed, and edited the manuscript. CM and JL drafted the manuscript. YC, XS, and MY performed the bioinformatic analysis. All authors contributed to the article and approved the submitted version.

Acknowledgments

The original data in our study were retrieved from the GEO database (https://www.ncbi.nlm.nih.gov/geo) and we gratefully acknowledge the contributions from public available databases.

Conflict of interest

The 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.2023.1161068/full#supplementary-material

References

Ballard, C., Gauthier, S., Corbett, A., Brayne, C., Aarsland, D., and Jones, E. (2011). Alzheimer's disease. *Lancet* 377, 1019–1031. doi: 10.1016/S0140-6736(10)61349-9

Cao, W., and Zheng, H. (2018). Peripheral immune system in aging and Alzheimer's disease. *Mol. Neurodegener.* 13:51. doi: 10.1186/s13024-018-0284-2

Chen, J., Song, Y.-W., Liang, G.-Z., Zhang, Z.-J., Wen, X.-F., Li, R.-B., et al. (2022). A novel m7G-related gene signature predicts the prognosis of colon cancer. *Cancers* 14:5527. doi: 10.3390/cancers14225527

Cheng, W., Gao, A., Lin, H., and Zhang, W. (2022). Novel roles of METTL1/WDR4 in tumor via m7G methylation. *Mol. Ther. Oncolytics* 26, 27–34. doi: 10.1016/j.omto.2022.05.009

Cui, L., Ma, R., Cai, J., Guo, C., Chen, Z., Yao, L., et al. (2022). RNA modifications: importance in immune cell biology and related diseases. *Signal Transduct. Target. Ther.* 7:334. doi: 10.1038/s41392-022-01175-9

Dai, Z., Liu, H., Liao, J., Huang, C., Ren, X., Zhu, W., et al. (2021). N7-Methylguanosine tRNA modification enhances oncogenic mRNA translation and promotes intrahepatic cholangiocarcinoma progression. *Mol. Cell* 81, 3339.e8–3355.e8. doi: 10.1016/j.molcel.2021.07.003

Gentles, A. J., Newman, A. M., Liu, C. L., Bratman, S. V., Feng, W., Kim, D., et al. (2015). The prognostic landscape of genes and infiltrating immune cells across human cancers. *Nat. Med.* 21, 938–945. doi: 10.1038/nm.3909

Gold, C., and Sollich, P. (2003). Model selection for support vector machine classification. *Neurocomputing* 55, 221–249. doi: 10.1016/S0925-2312(03)00375-8

Jevtic, S., Sengar, A. S., Salter, M. W., and McLaurin, J. (2017). The role of the immune system in Alzheimer disease: etiology and treatment. *Ageing Res. Rev.* 40, 84–94. doi: 10.1016/j.arr.2017.08.005

Kenny, A., Jiménez-Mateos, E. M., Zea-Sevilla, M. A., Rábano, A., Gili-Manzanaro, P., Prehn, J. H. M., et al. (2019). Proteins and microRNAs are differentially expressed in tear fluid from patients with Alzheimer's disease. *Sci. Rep.* 9:15437. doi: 10.1038/s41598-019-51837-y

Khan, A. A., Huang, H., Zhao, Y., Li, H., Pan, R., Wang, S., et al. (2022). WBSCR22 and TRMT112 synergistically suppress cell proliferation, invasion and tumorigenesis in pancreatic cancer via transcriptional regulation of ISG15. *Int. J. Oncol.* 60:24. doi: 10.3892/ijo.2022.5314

Li, B., Zhao, R., Qiu, W., Pan, Z., Zhao, S., Qi, Y., et al. (2022). The N6-methyladenosine-mediated lncRNA WEE2-AS1 promotes glioblastoma progression by stabilizing RPN2. *Theranostics* 12, 6363–6379. doi: 10.7150/thno.74600

Li, D.-X., Feng, D.-C., Wang, X.-M., Wu, R.-C., Zhu, W.-Z., Chen, K., et al. (2023). M7G-related molecular subtypes can predict the prognosis and correlate with immunotherapy and chemotherapy responses in bladder cancer patients. *Eur. J. Med. Res.* 28:55. doi: 10.1186/s40001-023-01012-x

Li, X., An, W.-L., Alafuzoff, I., Soininen, H., Winblad, B., and Pei, J.-J. (2004). Phosphorylated eukaryotic translation factor 4E is elevated in Alzheimer brain. *Neuroreport* 15, 2237–2240. doi: 10.1097/00001756-200410050-00019

Lin, C., Xu, C., Zhou, Y., Chen, A., and Jin, B. (2022). Identification of biomarkers related to M2 macrophage infiltration in Alzheimer's disease. *Cells* 11:2365. doi: 10.3390/cells11152365

Liu, M., Li, F., Yan, H., Wang, K., Ma, Y., Shen, L., et al. (2020). A multi-model deep convolutional neural network for automatic hippocampus segmentation and classification in Alzheimer's disease. *NeuroImage* 208:116459. doi: 10.1016/j.neuroimage.2019.116459

Lu, L., Zheng, J., Liu, B., Wu, H., Huang, J., Wu, L., et al. (2022). The m7G modification level and immune infiltration characteristics in patients with COVID-19. *J. Multidiscip. Healthc.* 15, 2461–2472. doi: 10.2147/JMDH.S385050

Luo, Y., Yao, Y., Wu, P., Zi, X., Sun, N., and He, J. (2022). The potential role of N7-methylguanosine (m7G) in cancer. *J. Hematol. Oncol.* 15:63. doi: 10.1186/s13045-022-01285-5

Maimaiti, A., Feng, Z., Liu, Y., Turhon, M., Xie, Z., Baihetiyaer, Y., et al. (2023). N7-methylguanosin regulators-mediated methylation modification patterns and characterization of the immune microenvironment in lower-grade glioma. *Eur. J. Med. Res.* 28:144. doi: 10.1186/s40001-023-01108-4

Mangialasche, F., Solomon, A., Winblad, B., Mecocci, P., and Kivipelto, M. (2010). Alzheimer's disease: clinical trials and drug development. *Lancet Neurol.* 9, 702–716. doi: 10.1016/S1474-4422(10)70119-8

Nelder, J. A., and Wedderburn, R. W. M. (1972). Generalized linear models. *J. R. Stat. Soc. Ser. A* 135, 370–384. doi: 10.2307/2344614

Pérez, R. F., Alba-Linares, J. J., Tejedor, J. R., Fernández, A. F., Calero, M., Román-Domínguez, A., et al. (2022). Blood DNA methylation patterns in older adults with evolving dementia. *J. Gerontol. Ser. A Biol. Sci. Med. Sci.* 77, 1743–1749. doi: 10.1093/gerona/glac068

Qin, H., Sheng, W., Weng, J., Li, G., Chen, Y., Zhu, Y., et al. (2023). Identification and verification of m7G-related genes as biomarkers for prognosis of sarcoma. *Front. Genet.* 14:1101683. doi: 10.3389/fgene.2023.1101683

Qiu, S., Miller, M. I., Joshi, P. S., Lee, J. C., Xue, C., Ni, Y., et al. (2022). Multimodal deep learning for Alzheimer's disease dementia assessment. *Nat. Commun.* 13:3404. doi: 10.1038/s41467-022-31037-5

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

Romeo, L., and Frontoni, E. (2022). A unified hierarchical XGBoost model for classifying priorities for COVID-19 vaccination campaign. *Pattern Recogn.* 121:108197. doi: 10.1016/j.patcog.2021.108197

Suresh, J., Khor, I. W., Kaur, P., Heng, H. L., Torta, F., Dawe, G. S., et al. (2021). Shared signaling pathways in Alzheimer's and metabolic disease may point to new treatment approaches. *FEBS J.* 288, 3855–3873. doi: 10.1111/febs.15540

Tian, Q.-H., Zhang, M.-F., Zeng, J.-S., Luo, R.-G., Wen, Y., Chen, J., et al. (2019). METTL1 overexpression is correlated with poor prognosis and promotes hepatocellular carcinoma via PTEN. *J. Mol. Med.* 97, 1535–1545. doi: 10.1007/s00109-019-01830-9

Wang, B., Fu, C., Wei, Y., Xu, B., Yang, R., Li, C., et al. (2022). Ferroptosis-related biomarkers for Alzheimer's disease: identification by bioinformatic analysis in hippocampus. *Front. Cell. Neurosci.* 16:1023947. doi: 10.3389/fncel.2022.1023947

Wei, C., Zhang, X., Peng, D., Zhang, X., Guo, H., Lu, Y., et al. (2022). LncRNA HOXA11-AS promotes glioma malignant phenotypes and reduces its sensitivity to ROS via Tpl2-MEK1/2-ERK1/2 pathway. *Cell Death Dis.* 13:942. doi: 10.1038/s41419-022-05393-5

Wu, X., Li, C., Wang, Z., Zhang, Y., Liu, S., Chen, S., et al. (2022). A bioinformatic analysis study of m7G regulator-mediated methylation modification patterns and tumor microenvironment infiltration in glioblastoma. *BMC Cancer* 22:729. doi: 10.1186/s12885-022-09791-y

Xia, P., Zhang, H., Xu, K., Jiang, X., Gao, M., Wang, G., et al. (2021). MYC-targeted WDR4 promotes proliferation, metastasis, and sorafenib resistance by inducing CCNB1 translation in hepatocellular carcinoma. *Cell Death Dis.* 12:691. doi: 10.1038/s41419-021-03973-5

Xia, X., Wang, Y., and Zheng, J. C. (2023). Internal m7G methylation: a novel epitranscriptomic contributor in brain development and diseases. *Mol. Ther. Nucleic Acids* 31, 295–308. doi: 10.1016/j.omtn.2023.01.003

Zhang, M., Song, J., Yuan, W., Zhang, W., and Sun, Z. (2021). Roles of RNA methylation on tumor immunity and clinical implications. *Fronti. Immunol.* 12:641507. doi: 10.3389/fimmu.2021.641507

Zhang, T., Liu, N., Wei, W., Zhang, Z., and Li, H. (2021). Integrated analysis of weighted gene coexpression network analysis identifying six genes as novel biomarkers for Alzheimer's disease. *Oxid. Med. Cell. Longev.* 2021:9918498. doi: 10.1155/2021/9918498

Zhong, C., Tao, B., Li, X., Xiang, W., Peng, L., Peng, T., et al. (2022). HOXA-AS2 contributes to regulatory T cell proliferation and immune tolerance in glioma through the miR-302a/KDM2A/JAG1 axis. *Cell Death Dis.* 13:160. doi: 10.1038/s41419-021-04471-4