Observational Study

Medicine

OPEN

m6A-related genes and their role in Parkinson's disease

Insights from machine learning and consensus clustering

Jing Yan, MD[a], Zhengyan Wang, MD[a], Yunqiang Li, MD[a], Ruien Li, MD[a], Ke Xiang, MD[b,*]

Abstract

Parkinson disease (PD) is a chronic neurological disorder primarily characterized by a deficiency of dopamine in the brain. In recent years, numerous studies have highlighted the substantial influence of RNA N6-methyladenosine (m6A) regulators on various biological processes. Nevertheless, the specific contribution of m6A-related genes to the development and progression of PD remains uncertain. In this study, we performed a differential analysis of the GSE8397 dataset in the Gene Expression Omnibus database and selected important m6A-related genes. Candidate m6A-related genes were then screened using a random forest model to predict the risk of PD. A nomogram model was built based on the candidate m6A-related genes. By employing a consensus clustering method, PD was divided into different m6A clusters based on the selected significant m6A-related genes. Finally, we performed immune cell infiltration analysis to explore the immune infiltration between different clusters. We performed a differential analysis of the GSE8397 dataset in the Gene Expression Omnibus database and selected 11 important m6A-related genes. Four candidate m6A-related genes (YTH Domain Containing 2, heterogeneous nuclear ribonucleoprotein C, leucine-rich pentatricopeptide repeat motif containing protein and insulin-like growth factor binding protein-3) were then screened using a random forest model to predict the risk of PD. A nomogram model was built based on the 4 candidate m6A-related genes. The decision curve analysis indicated that patients can benefit from the nomogram model. By employing a consensus clustering method, PD was divided into 2 m6A clusters (cluster A and cluster B) based on the selected significant m6A-related genes. The immune cell infiltration analysis revealed that cluster A and cluster B exhibit distinct immune phenotypes. In conclusion, m6A-related genes play a significant role in the development of PD and our study on m6A clustering may potentially guide personalized treatment strategies for PD in the future.

Abbreviations: AD = Alzheimer disease, AUC = area under the curve, DCA = decision curve analysis, DEGs = differentially expressed genes, ELAVL1 = ELAV Like RNA Binding Protein 1, GEO= Gene Expression Omnibus, HNRNPA2B1 = heterogeneous nuclear ribonucleoprotein A2/B1, HNRNPC = heterogeneous nuclear ribonucleoprotein C, IGFBP3 = insulin-like growth factor binding protein-3, KEGG = Kyoto Encyclopedia of Genes and Genomes, LRPPRC = leucine-rich pentatricopeptide repeat motif containing protein, METTL3 = methyltransferase like 3, m6A= N6-methyladenosine, PD = Parkinson disease, RF = random forest, ROC = receiver operating characteristic, ssGSEA = single-sample gene set enrichment analysis, SVM = support vector machine, YTHDC1 = YTH Domain Containing 1, YTHDC2 = YTH Domain Containing 2.

Keywords: consensus clustering, m6A methylation, m6A-related genes, machine learning model, Parkinson disease (PD)

1. Introduction

Parkinson disease (PD), also referred to as "shaking palsy," is a prevalent and progressive neurodegenerative disorder that affects the central nervous system. Its clinical presentation encompasses a spectrum of symptoms, including bradykinesia (slowness of movement), resting tremors, muscular rigidity, postural and gait abnormalities, and motor impairments.[1,2] PD currently ranks as the second most common neurodegenerative disease worldwide, second only to Alzheimer disease (AD). However, despite its prevalence, the precise etiology and pathogenesis of PD remain elusive.[3] Current research suggested that a combination of genetic factors, environmental influences, immune responses, inflammation, oxidative stress, and other factors collectively contribute to the degeneration and loss of dopaminergic neurons in the substantia nigra.[4] The primary pathological features of PD manifest as the gradual degeneration and loss of dopaminergic neurons in the midbrain's substantia nigra, resulting in impaired

The authors have no funding and conflicts of interest to disclose.

The datasets generated during and/or analyzed during the current study are publicly available.

All data are from public databases (GEO database), therefore, ethics committee approval is not required.

[a] *Changchun University of Chinese Medicine, Changchun, Jilin Province, China,* [b] *Jilin Provincial Academy of Chinese Medicine Sciences, Changchun, Jilin Province, China.*

** Correspondence: Ke Xiang, Jilin Provincial Academy of Chinese Medicine Sciences, Changchun 130117, Jilin Province, China (e-mail: xiangke1961@163.com).*

How to cite this article: Yan J, Wang Z, Li Y, Li R, Xiang K. m6A-related genes and their role in Parkinson's disease: Insights from machine learning and consensus clustering. Medicine 2024;103:45(e40484).

Received: 22 April 2024 / Received in final form: 4 June 2024 / Accepted: 24 October 2024

http://dx.doi.org/10.1097/MD.0000000000040484

dopamine production. Additionally, the remaining neurons exhibit the accumulation of protein aggregates known as Lewy bodies within their cytoplasm.[5,6] Recent studies have uncovered evidence implicating 7 rare highly penetrant monogenic alterations, namely SNCA, LRRK2, VPS35, PRKN, PINK1, DJ-1, and GBA, as strongly associated with familial cases of typical PD. Furthermore, mutations within 90 genetic loci have been linked to sporadic PD. Collectively, these genetic variations within the 90 loci account for 16% to 36% of the hereditary risk of developing PD.[7,8] Many of the genes associated with PD contribute to common biological pathways, and their interactions escalate the risk of PD onset and progression. Thus, early screening and effective prevention strategies targeted at high-risk individuals from a genetic standpoint hold immense potential for controlling the occurrence of PD.

RNA fulfills a dual role in genetic regulation and expression, serving as a crucial intermediary in the process of gene expression. This multifaceted process is accompanied by a diverse array of chemical modifications known as RNA modifications, predominantly observed in eukaryotic organisms.[9,10] To date, more than 160 types of RNA modifications have been identified, playing pivotal roles in gene transcription and posttranscriptional regulatory processes.[11] Among these modifications are N6-methyladenosine (m6A), 5-methylcytosine, N1-methyladenosine, and others, collectively contributing to the functional diversity and genetic information encoded within RNA.[12,13] Notably, m6A methylation stands as one of the most prevalent RNA modifications in eukaryotes.[14] Positioned predominantly in the 3' untranslated region (3' UTR) of mRNA and in close proximity to the stop codon, m6A methylation at the N6 position of adenosine plays a crucial regulatory role.[15,16] The m6A modification represents a significant epigenetic alteration necessitating the coordinated action of various regulatory proteins encoded by writers, erasers, and readers.[17] As a reversible mRNA modification, m6A exerts regulatory control over mRNA stability, splicing, transport, translation, localization, and can influence the higher-order structure of mRNA or disrupt protein–RNA interactions, thereby exerting influence over gene expression and impacting various biological processes. Notably, extensive research has illuminated the broad involvement of m6A in processes related to human growth, development, and metabolism.[18,19] Furthermore, the intricate association between m6A and various diseases, including tumors and obesity, has been established.[20,21] Investigation into mRNA methylation modifications has also unveiled connections to neurodevelopmental and neurodegenerative diseases.[22,23] Understanding the interplay between m6A RNA methylation and the immune microenvironment in PD offers insights into disease pathogenesis and potential therapeutic strategies. Targeting m6A regulators or immune-related pathways affected by m6A dysregulation may hold promise for mitigating neuroinflammation, preserving neuronal function, and slowing disease progression in PD patients. However, the precise role of m6A methylation in the context of PD remains incompletely understood. Thus, the objective of this study is to elucidate the correlation between the expression of m6A-related genes and PD, thereby unraveling the intricate association between m6A methylation and the pathogenesis of PD. This research has the potential to significantly contribute to the prevention and treatment of PD.

2. Materials and methods

2.1. Source and processing of data

The Gene Expression Omnibus (GEO) is an international public repository that collects and organizes high-throughput genomic data, such as microarray chips and next-generation sequencing data, uploaded by researchers worldwide.[24] In this study, we utilized the GEO database and downloaded the GSE8397 dataset, which consists of 18 control samples and 29 PD samples. The dataset was generated using the GPL96 platform, specifically the Affymetrix Human Genome U133A Array (HG-U133A).[25] To identify differentially expressed genes (DEGs) associated with m6A, we performed differential analysis using the "limma" package in R version 4.2.2. This analysis was conducted on the control group samples and PD samples. Finally, we validated our results using the GSE22491 dataset (controls: 8, PD: 10) and the GSE28894 dataset (controls: 59, PD: 55).[26] Since the data used in our study is entirely sourced from an open database, no additional ethical approval was required for our research.

2.2. The construction and selection of machine learning models

In order to predict the onset of PD, we constructed 2 machine learning models: random forest (RF) and support vector machine (SVM). RF is an ensemble learning method comprising multiple decision trees. The final prediction of RF is based on the average prediction results of all decision trees, making it a widely used and effective machine learning model.[27] On the other hand, SVM aims to find a hyperplane that effectively separates data points of different classes while maximizing the distance between the hyperplane and the nearest data points, known as support vectors. SVM is particularly useful for handling small-sized datasets with limited samples.[28] For our study, we employed the "RandomForest" package for building the RF model and the "kernlab" package for constructing the SVM model, using R 4.2.2. These packages provide robust and efficient tools for implementing the respective machine learning algorithms. The performance of both models was evaluated using various metrics, including the "reverse cumulative distribution of residuals," "residual boxplots," and receiver operating characteristic (ROC) curves. These evaluations help assess the accuracy and predictive capabilities of the models. Based on the comprehensive evaluation of the 2 machine learning models, we selected the best-performing model for further research and analysis.

2.3. The construction and analysis of the nomogram model

In our study, we employed the "rms" package in R 4.2.2 to construct a nomogram model based on the selected candidate m6A-related genes, aiming to predict the incidence of PD in patients. The nomogram model provides a visual representation of the predictive model, displaying the contribution and weightage of each variable. To evaluate the performance of the nomogram model, we employed various assessment techniques. Calibration curves were plotted to assess the agreement between the predicted probabilities from the model and the observed outcomes. This allows us to evaluate the calibration or accuracy of the model predictions. Decision curve analysis (DCA) was conducted to evaluate the clinical utility of the model. DCA helps assess the net benefits of using the model-based predictions compared to alternative strategies or clinical decisions. By examining the threshold probabilities, DCA enables us to determine the clinical impact and usefulness of the nomogram model. Additionally, clinical impact curves were plotted to visualize the potential impact of the model-based decisions on patient outcomes. These curves provide insights into the potential benefits of using the nomogram model in clinical decision-making. By utilizing these evaluation techniques, we aim to assess the performance, accuracy, and clinical utility of the nomogram model in predicting the incidence of PD, ultimately determining whether the model-based decisions are beneficial for patients.

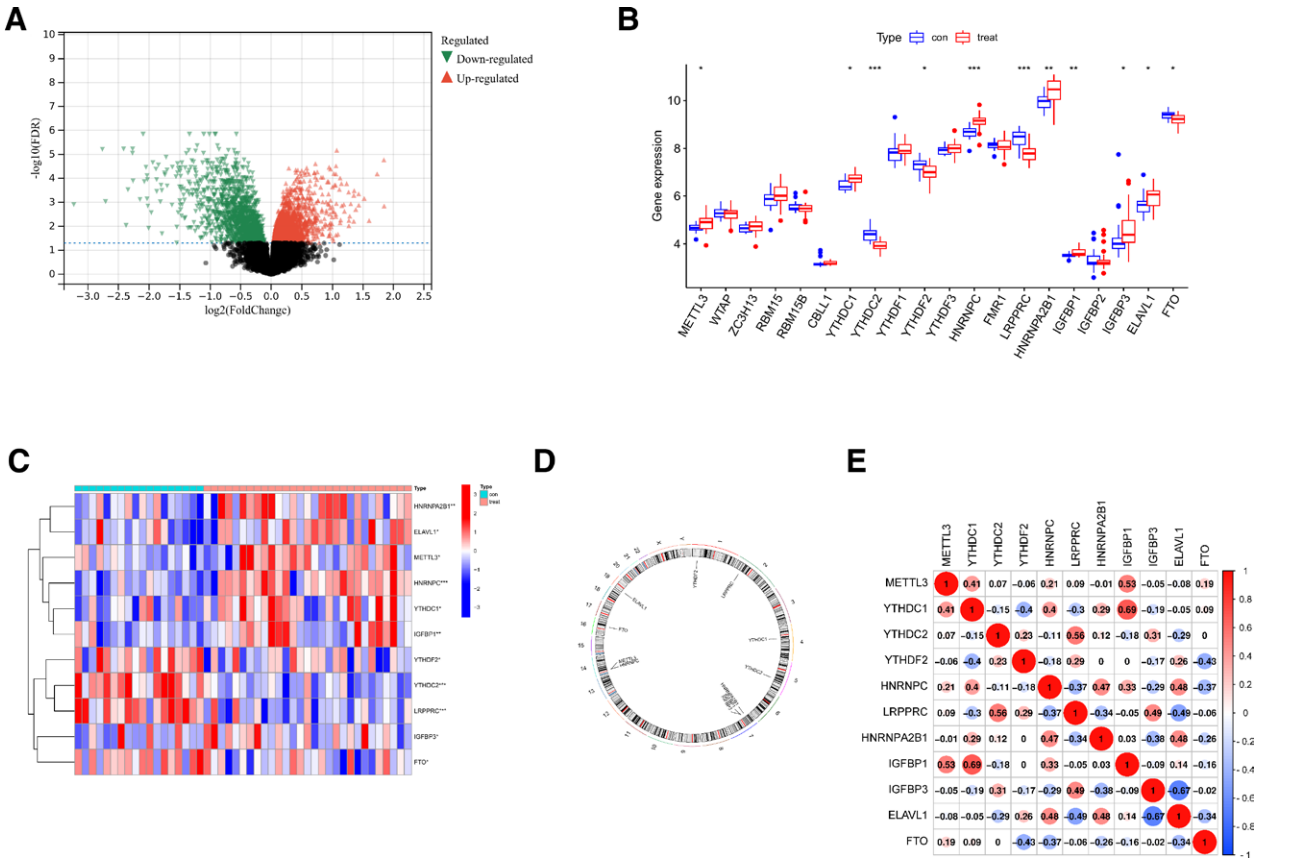

Figure 1. Distribution of RNA N6 methyl adenosine (m6A)-related genes in Parkinson disease (PD). (A) A volcanic map of the differential genes identified in GSE8397 (PD) dataset. (B) Boxplots of differential expression of m6A-related genes identified between control samples and PD samples. (C) Heatmap of the expression of 11 m6A-related genes in control and PD samples. (D) Location of the 11 m6A-related genes on the chromosome. (E) 11 m6A gene correlation analysis. *$P < .05$, **$P < .01$, and ***$P < .001$.

2.4. Identification and analysis of clusters based on 11 m6A genes

Consensus clustering is a widely used approach for analyzing disease clusters, whereby samples are grouped into distinct clusters based on various omics datasets. This method facilitates the identification of novel disease clusters and enables comparative analysis among different clusters.[29] In our study, we employed the "Consensus ClusterPlus" package in R 4.2.2 to perform consensus clustering based on m6A-related genes, with the objective of identifying different m6A clusters. This package provides a comprehensive framework for executing consensus clustering analysis. Subsequently, employing a significance threshold of $P < .01$, we utilized the "limma" package in R 4.2.2 to identify DEGs between the m6A clusters. The "limma" package is a powerful tool for differential gene expression analysis, allowing us to discern genes that exhibit significant expression differences between the identified clusters. Finally, to gain insights into the biological functions and pathways associated with the DEGs, we conducted gene ontology and Kyoto Encyclopedia of Genes and Genomes enrichment analysis. The Metascape database was employed for this purpose, as it provides a comprehensive platform for gene annotation and functional enrichment analysis. By utilizing these methods, we aim to unravel the distinctive characteristics and underlying biological processes associated with different m6A clusters, providing valuable insights into the molecular mechanisms involved in PD and paving the way for further research.

2.5. Immune infiltration analysis of m6A clusters

In our study, we employed single-sample gene set enrichment analysis (ssGSEA) to evaluate the abundance of immune cells in PD samples. The ssGSEA method enables the assessment of the enrichment scores of gene expression levels in each sample. To perform ssGSEA, we first obtained the gene expression levels of the samples and calculated their respective enrichment scores. These scores reflect the degree of enrichment or activation of specific gene sets within each sample. Next, we searched for the genes associated with immune cells in the input dataset and computed the sum of their corresponding enrichment scores. This summation provides an estimation of the abundance of immune cells in each individual sample. By utilizing ssGSEA, we aimed to gain insights into the relative abundance of immune cells within PD samples. This information can contribute to our understanding of the immune landscape and its potential implications in the pathogenesis and progression of PD. By employing these methods, we can further explore the intricate relationship between immune cell abundance and PD, potentially uncovering novel insights into the immunological mechanisms involved in the disease.[30]

3. Results

3.1. Analysis of m6A-related DEGs

In our study, we conducted differential expression analysis of genes between control and PD samples using the "limma" package in R 4.2.2 (Fig. 1A). The analysis

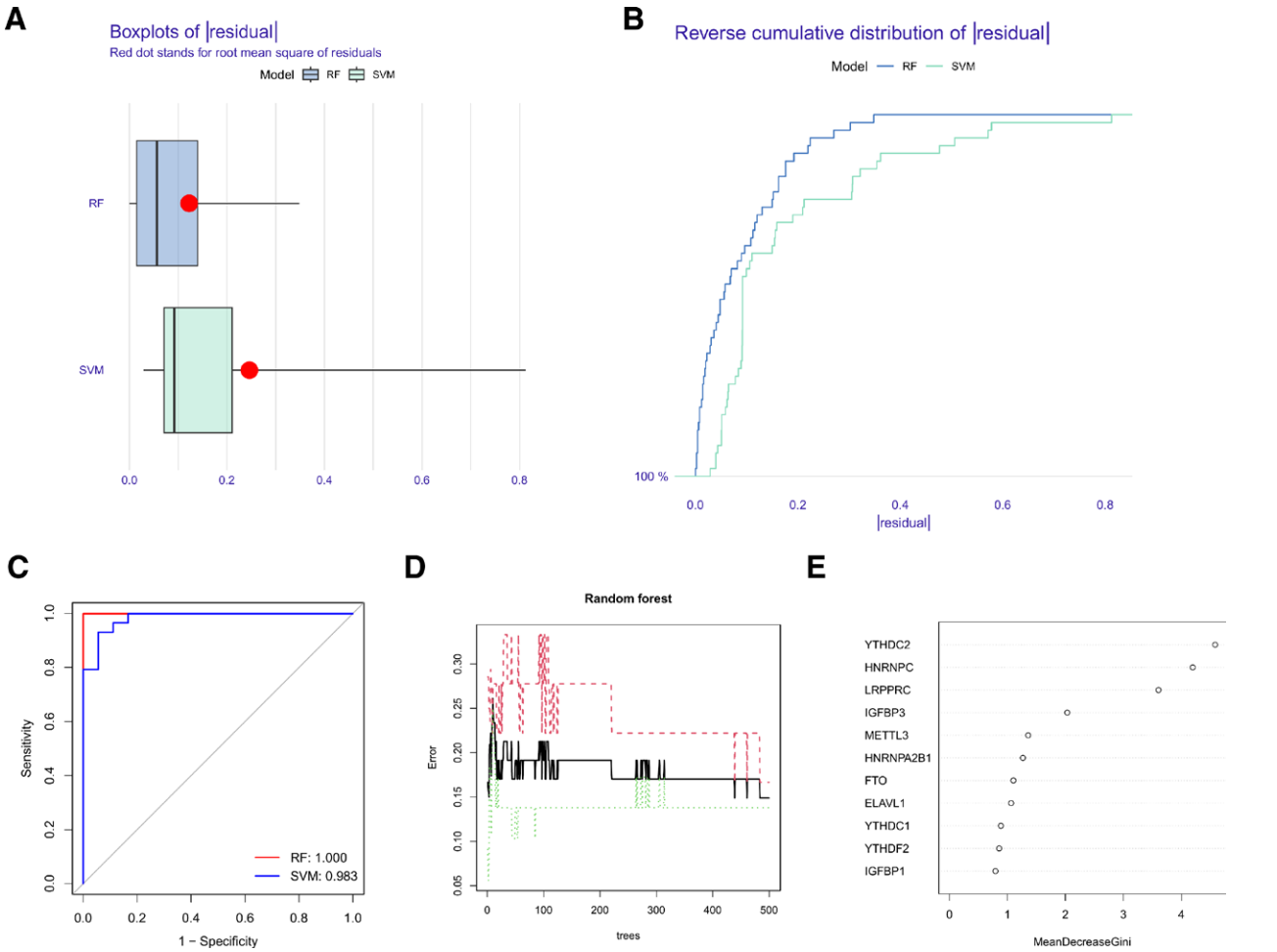

Figure 2. RF and SVM machine learning model construction. (A) Box plots of RF and SVM residuals to show the distribution of residuals for the RF and SVM models. (B) Inverse cumulative distribution of RF and SVM residuals to show the distribution of residuals for RF and SVM models. (C) ROC curves show the accuracy of RF and SVM models. (D) Random forest tree results. (E). Importance scores for disease characterizing genes. RF = random forest, ROC = receiver operating characteristic, SVM = support vector machine.

revealed significant differential expression of 26 m6A regulators. Among them, 11 m6A-related genes exhibited significant differential expression. These genes included 1 writer (methyltransferase like 3 [METTL3]), 1 eraser (fat mass and obesity-associated protein), and 9 readers (YTH Domain Containing 1 [YTHDC1], YTH Domain Containing 2 [YTHDC2], YTH N6-methyladenosine RNA binding protein 2, heterogeneous nuclear ribonucleoprotein C [HNRNPC], leucine-rich pentatricopeptide repeat motif containing protein [LRPPRC], heterogeneous nuclear ribonucleoprotein A2/B1 [HNRNPA2B1], insulin-like growth factor binding protein-3 [IGFBP1], IGFBP3, ELAV Like RNA Binding Protein 1 [ELAVL1]) (Fig. 1B). To visualize the expression patterns of these 11 m6A-related genes, we generated a heatmap (Fig. 1C). The heatmap highlighted the differential expression levels of these genes between control and PD samples. It was observed that METTL3, YTHDC1, HNRNPC, HNRNPA2B1, IGFBP1, IGFBP3, and ELAVL1 were upregulated in PD samples, while YTHDC2, YTH N6-methyladenosine RNA binding protein 2, LRPPRC, and fat mass and obesity-associated protein were downregulated in PD samples. To gain insights into the genomic distribution of the 11 m6A regulators, we utilized the "RCircos" package to generate a visualization of their chromosomal locations (Fig. 1D). Furthermore, we employed linear regression analysis to explore the correlations between m6A genes in PD. The analysis revealed strong positive correlations between the expression levels of IGFBP1 and YTHDC1, as well as METTL3 in PD samples. Additionally, a strong positive correlation was observed between YTHDC1 and METTL3, as well as RNA binding motif protein 15B3. On the other hand, the expression of YTHDC1 showed a negative correlation with YTHDC2. ELAVL1 exhibited a positive correlation with HNRNPC but a negative correlation with LRPPRC (Fig. 1E). These findings indicate that differentially expressed m6A-related genes in PD samples exhibit distinct correlations with each other. These results shed light on the differential expression patterns and interrelationships of m6A-related genes in the context of PD, providing valuable insights into the potential roles of these genes in disease pathogenesis.

3.2. Construction and screening of machine learning models

We constructed 2 machine learning models, RF and SVM, utilizing 11 differentially expressed m6A-related genes to predict the onset of PD. Upon examining the "Boxplots of Residuals" (Fig. 2A) and the "Reverse Cumulative Distribution of Residuals" (Fig. 2B), it is evident that the RF model exhibits the smallest residuals, signifying its superior performance.

Transitioning to the evaluation phase, the ROC curve was employed to assess the accuracy of these models. Notably, the RF model achieved an impeccable area under the curve (AUC) score of 1.0, outshining the SVM model, which secured an AUC of 0.983 (Fig. 2C). This further corroborates the RF model's heightened accuracy. In the final stage, we visualized the RF model (Fig. 2D) and identified genes with importance scores exceeding 2, namely YTHDC2, HNRNPC, LRPPRC, and IGFBP3, as prime candidates (Fig. 2E).

3.3. Construction of nomogram model

We proceeded to construct a nomogram model, utilizing the 4 identified m6A-related genes (YTHDC2, HNRNPC, LRPPRC, and IGFBP3), to predict the prevalence of PD. This was achieved through the "rms" package in R version 4.2.2 (Fig. 3A). Moving forward, the calibration curve was plotted and it revealed an outstanding predictive performance of the nomogram model (Fig. 3B). This is a pivotal step as it validates the reliability of the predictions made by the model. Subsequently, we conducted a DCA, the results of which suggest that employing the nomogram model for decision-making could be advantageous for patients diagnosed with PD (Fig. 3C). Lastly, the clinical impact curve was drawn, which further attests to the significant predictive capability of the nomogram model (Fig. 3D).

3.4. Identification of clusters based on 11 related m6A genes

Utilizing the "ConsensusClusterPlus" package in R version 4.2.2, consensus clustering techniques were employed to identify distinct m6A clusters based on 11 significant m6A-related genes, eventually identifying 2 m6A clusters, namely cluster A and cluster B (Fig. 4A–D). The cumulative distribution function curve exhibited changes within the consensus index's minimum range from 0.2 to 0.6 (Fig. 4E). Cluster A comprises 16 PD samples, whereas cluster B encompasses 13 PD samples. Heatmaps and box plots were subsequently generated to display the differential expression levels of the 11 critical m6A-related genes between the 2 clusters. IGFBP3 exhibits elevated expression levels in cluster A compared to cluster B, while HNRNPA2B1 and ELAVL1 display higher expression levels in cluster B than in cluster A (Fig. 4F and G). Principal component analysis reveals that the 11 critical m6A-related genes can distinctly differentiate the 2 m6A clusters (Fig. 4H). A total of 35 DEGs associated with m6A were chosen between the 2 m6A clusters. To explore the potential roles of these DEGs in PD, we conducted Kyoto Encyclopedia of Genes and Genomes enrichment analyses (Fig. 4I).

3.5. Immuno-infiltration analysis of m6A clusters

We applied ssGSEA to calculate the abundance of immune cells in PD samples and evaluated the correlation between 11

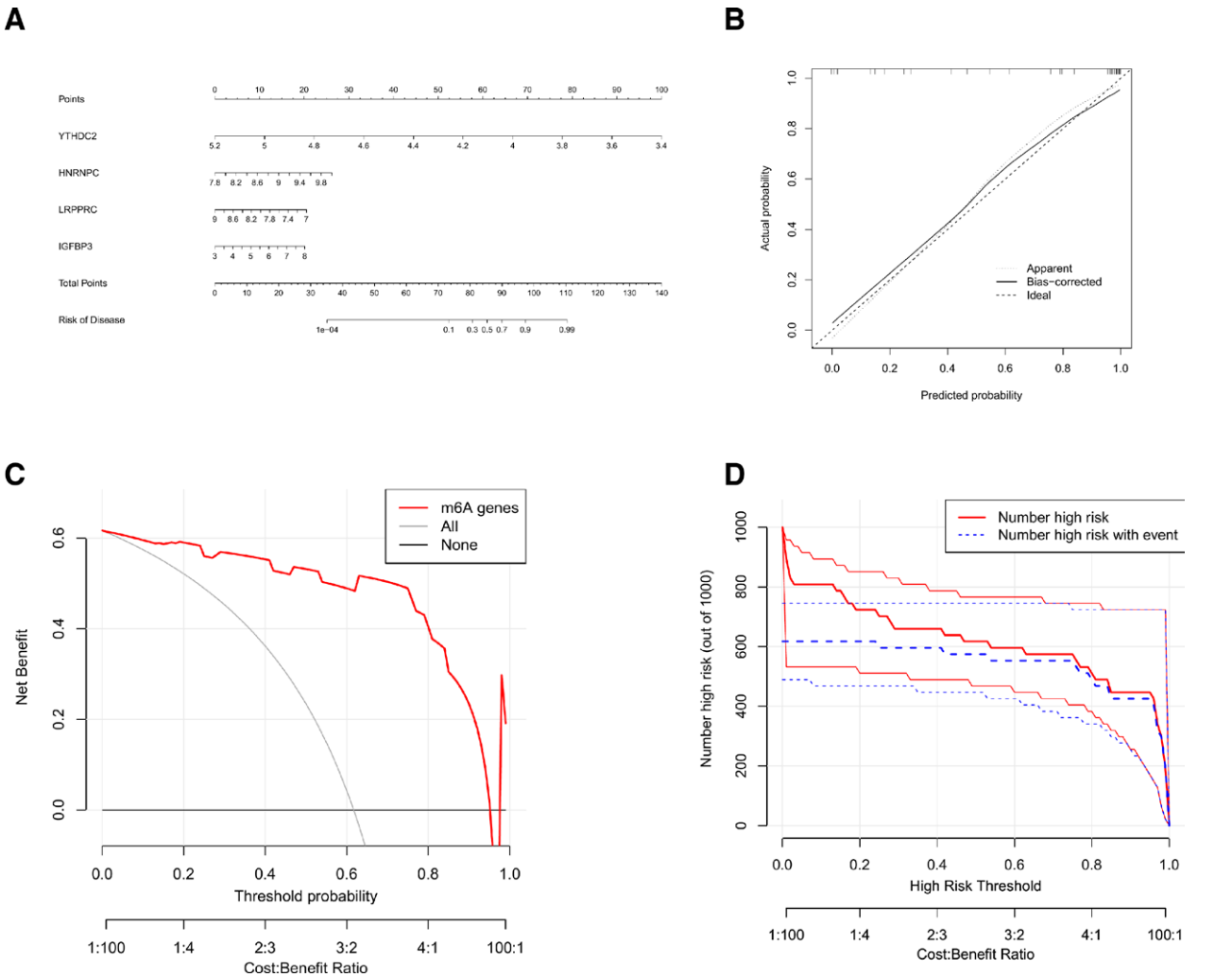

Figure 3. Construction of nomogram model. (A) Construction of nomogram model based on 4 candidate m6A-related genes. (B) Construction of calibration curve of the nomogram model. (C) Construction of DCA of the nomogram model. (D) Clinical impact curves to assess the clinical impact of nomogram models. DCA = decision curve analysis.

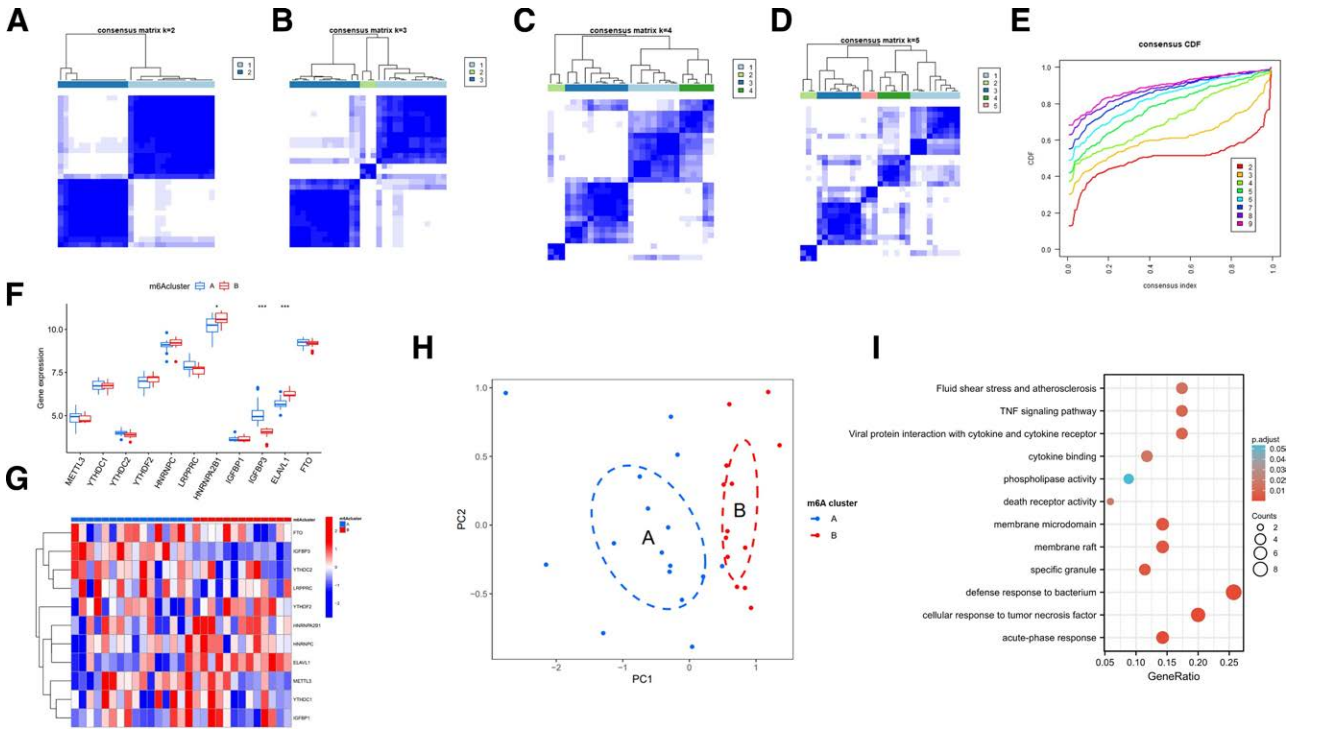

Figure 4. Consistent clustering of 11 m6A-related genes in Parkinson disease. (A) Consensus clustering matrix when k = 2. (B) Consensus clustering matrix when k = 3. (C) Consensus clustering matrix when k = 4. (D) Consensus clustering matrix when k = 5. (E) Representative CDF curves. (F) Box plots of differential expression of 11 m6A-related genes in cluster A and cluster B. (G) Heatmap of the expression of 11 m6A-related genes in cluster A and cluster B. (H) Principal component analysis of cluster A and cluster B. (I) KEGG enrichment analysis of 35 DEGs. *P < .05, **P < .01, and ***P < .001. CDF = cumulative distribution function, KEGG = Kyoto Encyclopedia of Genes and Genomes, m6A = N6-methyladenosine.

important m6A-related genes and immune cells. We found that activated CD4 T cells, activated dendritic cells, CD56 bright natural killer cells, gamma delta T cells, MDSCs (myeloid-derived suppressor cells), macrophages, natural killer T cells, natural killer cells, neutrophils, plasmacytoid dendritic cells, regulatory T cells, T follicular helper cells, and type 1 T helper cells have significantly different expression levels between cluster A and cluster B (Fig. 5A). Subsequently, we explored the correlation between m6A genes and immune cells. The results showed that HNRNPA2B1 is positively correlated with activated dendritic cells, regulatory T cells, and activated CD4 T cells, while it is negatively correlated with monocytes (Fig. 5B and C).

3.6. Identification of 2 genetic clusters and immune cell infiltration analysis

By employing the consensus clustering technique based on 35 m6A-related DEGs, PD samples were categorized into distinct clusters. We observed that there are 2 disparate m6A gene clusters (namely gene cluster A and gene cluster B), which aligns with the classification of m6A clusters (as depicted in Fig. 6A–D). The expression levels of the 35 m6A-related DEGs in gene cluster A and gene cluster B are illustrated in Figure 6E. Figure 6F and G demonstrates that the variations in expression levels of the 11 critical m6A genes and immune cell infiltration between gene cluster A and gene cluster B are akin to those in the m6A clusters. This lends further credence to the precision of our categorization through the consensus clustering approach. Finally, we validated our RF model using external independent datasets (GSE22491 and GSE28894). The constructed ROC results showed an AUC value of 0.582 for GSE22491 and an AUC value of 0.615 for GSE28894. All of these results demonstrated the good diagnostic efficacy of the RF model, as well as the reliability and accuracy of our findings (Fig. 7A and B).

4. Discussion

PD stands as a neurodegenerative ailment typified by the demise or impairment of neurons, subsequently leading to compromised motor function. Its etiology is thought to arise from a confluence of genetic and environmental factors.[31,32] Mounting evidence indicates the involvement of m6A-related genes in the biological mechanisms of diverse ailments.[33] Nevertheless, the precise implications of m6A-related genes in PD remain incompletely understood. Hence, this present investigation embarked upon differential expression analysis utilizing control and PD samples from the GSE8397 dataset, which successfully identified 11 pivotal m6A-related genes exhibiting disparate expression levels. Subsequent to this discovery, we devised a RF model founded upon these 11 differentially expressed m6A-related genes, with the aim of predicting the onset of PD. Moreover, utilizing the ranking system of feature gene importance scores from the RF model, we formulated a nomogram model centered around 4 promising candidate m6A-related genes (YTHDC2, HNRNPC, LRPPRC, and IGFBP3). By virtue of DCA, it was evident that the nomogram model held potential benefits for PD patients. Lastly, predicated on the aforementioned 11 m6A-related genes, we categorized the PD samples into 2 distinct clusters, namely cluster A and cluster B, whereupon we conducted immune infiltration analysis on these clusters. The ensuing results unearthed noteworthy disparities in various immune cells between cluster A and cluster B, strongly hinting at the existence of discrete immune phenotypes. In the context of PD, dysregulation of m6A modification has been observed in the brains of PD patients. This dysregulation affects the expression and activity of m6A writers, erasers, and readers, consequently impacting the stability, translation efficiency, and function of immune-related transcripts.[34] The immune microenvironment in PD is characterized by chronic neuroinflammation, marked by the activation of microglia and infiltration of peripheral immune cells into the brain parenchyma. Dysregulated immune responses contribute

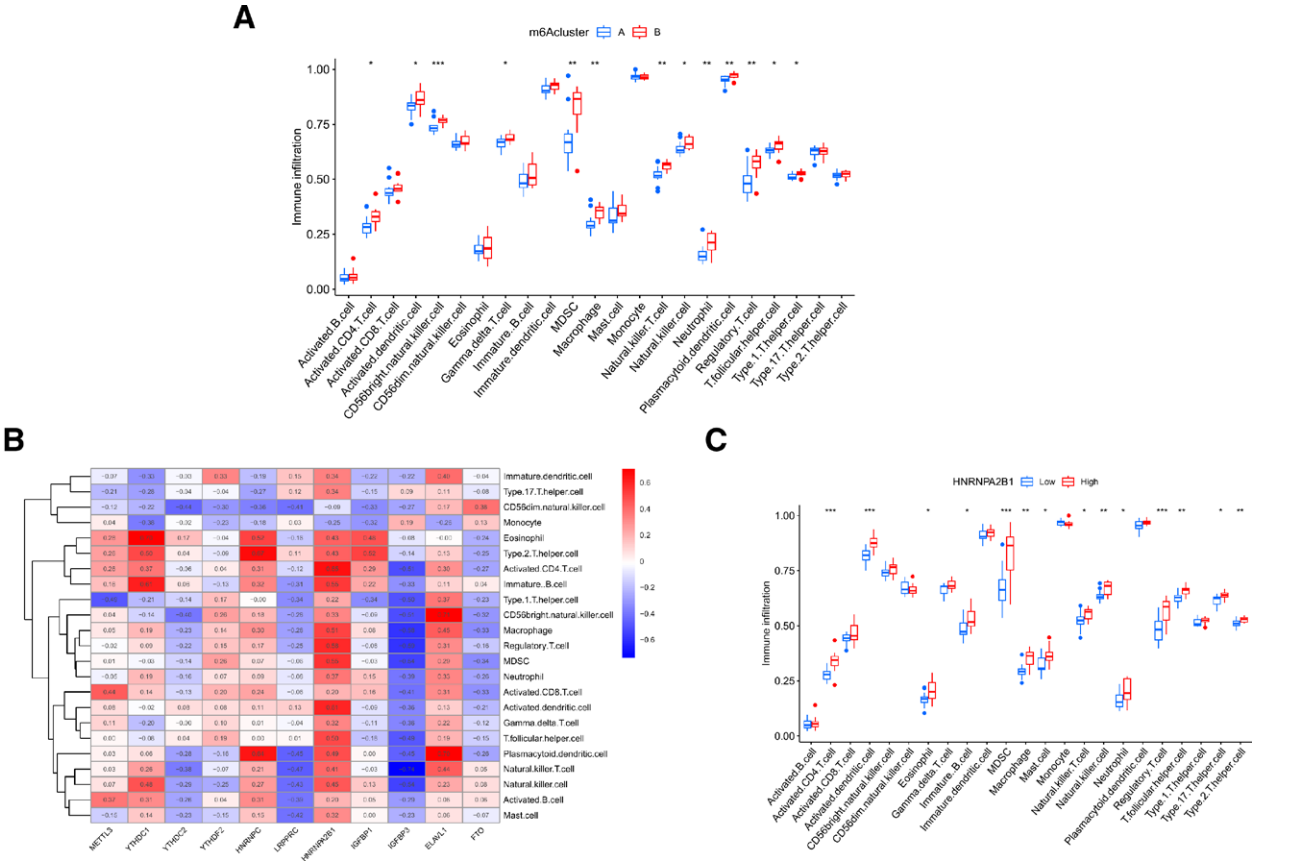

Figure 5. Analysis of immune cell infiltration by m6A subtypes. (A) Differential immune cell infiltration between cluster A and cluster B. (B) Correlation analysis between infiltrating immune cells and 11 m6A-related genes. (C) Difference in the abundance of infiltrating immune cells between high and HNRNPA2B1 expression groups. *$P < .05$, **$P < .01$, and ***$P < .001$. HNRNPA2B1 = heterogeneous nuclear ribonucleoprotein A2/B1, m6A = N6-methyladenosine.

to neuronal damage and disease progression.[35] Importantly, m6A RNA methylation influences the expression of immune-related genes and modulates immune responses. For instance, m6A modification can regulate the expression of cytokines, chemokines, and immune receptors, as well as modulate innate and adaptive immune cell functions.[36,37]

Intriguing research findings have illuminated the indispensability of specific binding proteins for the manifestation of the biological functions attributed to m6A modification. Among these, the YTH protein family emerges as a notable cohort, adept at discerning and selectively recognizing m6A-modified mRNA, thereby orchestrating its functional role.[38,39] Notably, YTH Domain Containing 2 (YTHDC2), the largest member of the YTH family, assumes a prominent position within the nucleus. Its principal undertaking revolves around augmenting the translation efficiency of downstream target genes.[40] As a significant m6A recognition protein and one of the m6A readers, YTHDC2 is primarily engaged in the recruitment of m6A-binding proteins. Notably, YTHDC2 exhibits a predilection for binding to noncoding RNAs, in addition to introns and intergenic regions.[22] Possessing adenosine triphosphate-dependent RNA helicase activity, YTHDC2 harbors 2 operative domains: Helicase N and Helicase C. The anchor protein repeat sequence interposed between these 2 domains facilitates the interaction with the ribonuclease XRN1, suggesting a putative role for YTHDC2 in regulating mRNA stability.[41,42]

Heterogeneous nuclear ribonucleoprotein C (HNRNPC) gene belongs to a subfamily of heterogeneously expressed ribonucleoproteins (hnRNPs) that are ubiquitously distributed throughout the cell. Intriguingly, m6A has been demonstrated to facilitate the interaction between HNRNPC and mRNA, mediated by a mechanism aptly termed the "m (6)A-switch," thereby exerting regulatory control over mRNA splicing events.[43] Current investigations have unveiled the involvement of HNRNPC in the malignant progression of diverse tumors, shedding light on its potential significance as a therapeutic target.[44–46] Furthermore, studies have indicated that augmented expression of HNRNPC can engender enhanced cellular proliferation and hinder apoptosis in PC2019 cells. Notably, it also serves to suppress the expression of inflammatory factors such as IFN-β, IL-2, and TNF-α. These intriguing observations posit the possibility of HNRNPC contributing to the pathogenesis of PD by impeding dopaminergic neuronal proliferation, promoting apoptosis, and eliciting immune-mediated inflammation.[37]

Leucine-rich pentatricopeptide repeat motif containing protein (LRPPRC), primarily situated on chromosome 2, assumes a pivotal role in the early onset of mitochondrial disorders, thus underscoring its significance in cellular homeostasis.[47,48] Extending beyond its involvement in mitochondrial diseases, LRPPRC has been implicated in an array of pathological conditions. Notably, in the realm of oncological investigations, LRPPRC has exhibited heightened expression across multiple cancer types, including but not limited to prostate cancer, colon cancer, gastric cancer, and lung cancer.[49–52] Furthermore, emerging studies have put forth the notion that heightened LRPPRC expression within the brain may disrupt mitochondrial adenosine triphosphate production, thereby influencing the onset and progression of neurodegenerative disorders such as AD and PD.[53]

Insulin-like growth factor binding protein-3 (IGFBP3), a prominent member of the IGFBP protein family, assumes a multifaceted role in cellular physiology. Research has elucidated that IGFBP3 exerts influence over cell proliferation and apoptosis via both IGF-dependent and IGF-independent signaling

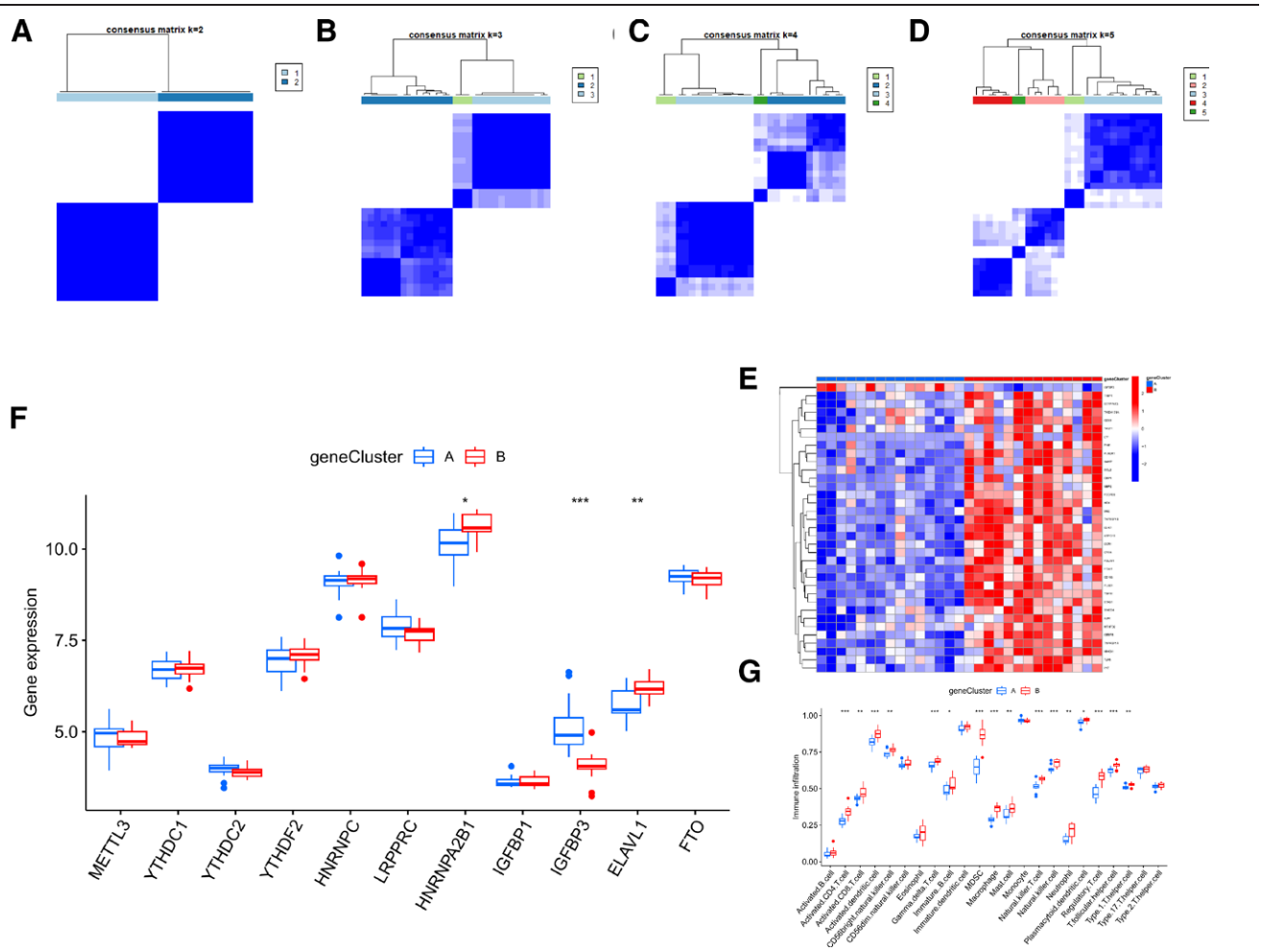

Figure 6. Consensus clustering of 35 m6A-related DEGs in Parkinson disease. (A–D) Consensus matrix of 35 m6A-related DEGs with K = 2–5. (E) Expression heatmap of 35 m6A-related DEGs in gene cluster A and gene cluster B. (F) Boxplot of differential expression of 11 m6A-related genes in gene cluster A and gene cluster B. (G) Differential analysis of immune cell infiltration of gene cluster A and gene cluster B. *$P < .05$, **$P < .01$, and **$P < .001$. DEGs = differentially expressed genes, m6A = N6-methyladenosine.

pathways.[54] Moreover, IGFBP3 has emerged as a key regulator of tumor cell growth, underscoring its significance in cancer biology.[55] Notably, studies have provided insights into the role of TRAIP, acting through the IGFBP3/AKT pathway, in promoting the degradation of KANK1, thus amplifying the invasive and proliferative capacities of osteosarcoma cells.[56] In the context of AD, investigations have uncovered the intriguing interplay between astrocyte-produced IGFBP3 and tau phosphorylation in neurons, thus imparting an impact on the progression of AD.[57] This study unveils the potential role of m6A-related genes in PD, enhancing our understanding of its pathogenic mechanisms. These genes likely play crucial roles in the development of PD, offering avenues for discovering novel therapeutic targets. By identifying these genes and elucidating their roles in PD, the study provides guidance for personalized treatment, aiding clinicians in assessing disease progression and devising more effective therapeutic strategies. Moreover, it sets a direction for future research, providing a framework for exploring the specific roles of these genes in PD and their interactions with other factors. This study furnishes invaluable insights into understanding the pathogenesis of PD, guiding clinical interventions, and inspiring future investigations, with the potential to substantially improve the health outcomes of patients.

Nevertheless, it is important to acknowledge the limitations of this study. Firstly, the absence of dedicated datasets specifically tailored to m6A-related genes within publicly available databases impeded our ability to independently validate the model. Consequently, the generalizability of our findings may be constrained. Moreover, the current comprehension surrounding the roles of the 4 identified genes (YTHDC2, HNRNPC, LRPPRC, and IGFBP3) in relation to PD remains limited, with the precise underlying mechanisms of their involvement in PD yet to be fully elucidated. In light of these limitations, future endeavors shall encompass conducting pertinent cell and animal experiments to corroborate the potential value of our model and shed light on the specific mechanisms underpinning PD pathogenesis. By doing so, we aim to further advance our understanding of this complex disorder.

5. Conclusion

Our study has achieved significant milestones by constructing a robust RF model and devising a nomogram model based on 4 promising m6A-related genes, facilitating accurate prediction of PD occurrence. Additionally, leveraging the insights gleaned from the 11 m6A-related genes, we have identified 2 distinct m6A clusters within PD, characterized by contrasting immune phenotypes. The culmination of these findings holds the potential to inspire new avenues of research, offering fresh perspectives and guiding future investigations pertaining to PD.

Acknowledgments

Thanks to the original data providers of the GEO database and Open GWAS database.

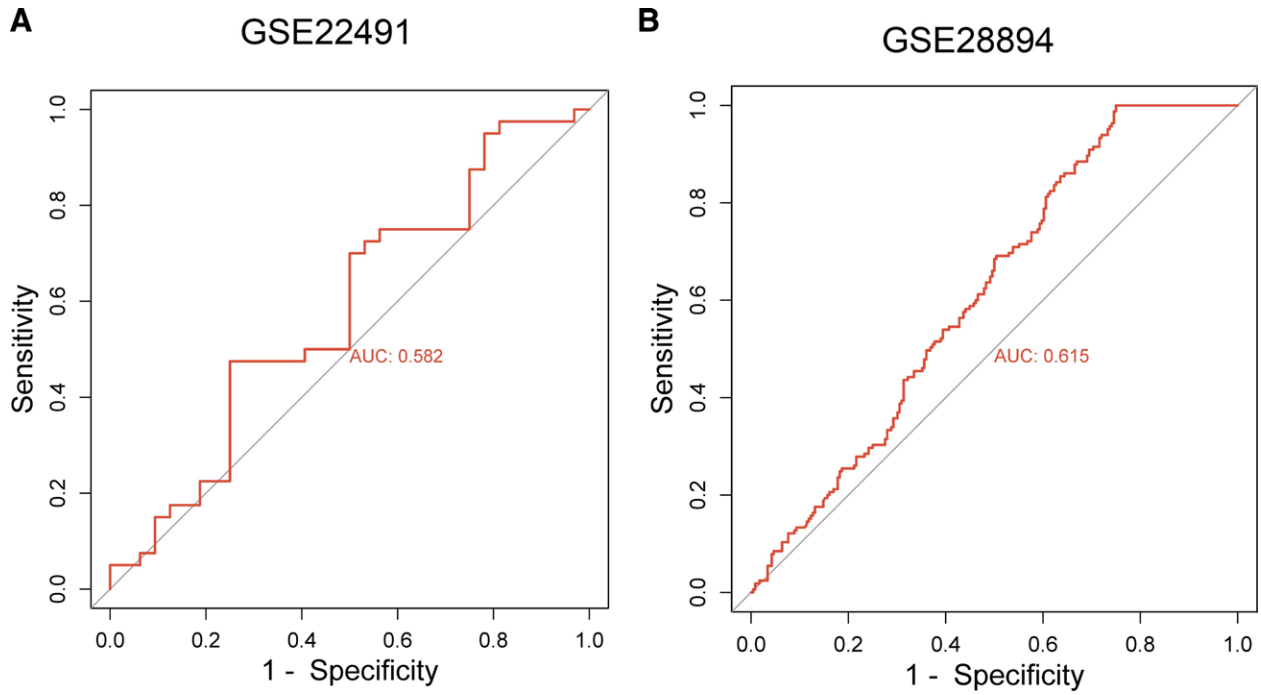

Figure 7. Independent external dataset validation results. (A) ROC results of GSE22491. (E) ROC results of GSE28894. ROC = receiver operating characteristic.

Author contributions

Conceptualization: Jing Yan, Zhengyan Wang, Yunqiang Li, Ruien Li, Ke Xiang.
Data curation: Jing Yan, Zhengyan Wang, Yunqiang Li, Ruien Li, Ke Xiang.
Formal analysis: Jing Yan, Ke Xiang.
Project administration: Ke Xiang.
Resources: Ke Xiang.
Visualization: Ke Xiang.
Writing – original draft: Jing Yan, Ke Xiang.
Writing – review & editing: Jing Yan, Zhengyan Wang, Yunqiang Li, Ke Xiang.

References

[1] Bloem BR, Okun MS, Klein C. Parkinson's disease. Lancet. 2021;397:2284–303.
[2] Jankovic J, Tan EK. Parkinson's disease: etiopathogenesis and treatment. J Neurol Neurosurg Psychiatry. 2020;91:795–808.
[3] Chen Z, Li G, Liu J. Autonomic dysfunction in Parkinson's disease: Implications for pathophysiology, diagnosis, and treatment. Neurobiol Dis. 2020;134:104700.
[4] Flores-Dorantes MT, Díaz-López YE, Gutiérrez-Aguilar R. Environment and gene association with obesity and their impact on neurodegenerative and neurodevelopmental diseases. Front Neurosci. 2020;14:863.
[5] Farrer MJ. Genetics of Parkinson disease: paradigm shifts and future prospects. Nat Rev Genet. 2006;7:306–18.
[6] Gray F. [The neuropathology of Parkinson syndrome]. Rev Neurol (Paris). 1988;144:229–48.
[7] Lunati A, Lesage S, Brice A. The genetic landscape of Parkinson's disease. Rev Neurol (Paris). 2018;174:628–43.
[8] Nalls MA, Blauwendraat C, Vallerga CL, et al. Identification of novel risk loci, causal insights, and heritable risk for Parkinson's disease: a meta-analysis of genome-wide association studies. Lancet Neurol. 2019;18:1091–102.
[9] Jonkhout N, Tran J, Smith MA, Schonrock N, Mattick JS, Novoa EM. The RNA modification landscape in human disease. RNA. 2017;23:1754–69.
[10] Barbieri I, Kouzarides T. Role of RNA modifications in cancer. Nat Rev Cancer. 2020;20:303–22.
[11] Boccaletto P, Stefaniak F, Ray A, et al. MODOMICS: a database of RNA modification pathways. 2021 update. Nucleic Acids Res. 2022;50:D231–5.
[12] Dominissini D, Nachtergaele S, Moshitch-Moshkovitz S, et al. The dynamic N(1)-methyladenosine methylome in eukaryotic messenger RNA. Nature. 2016;530:441–6.
[13] Yang X, Yang Y, Sun BF, et al. 5-methylcytosine promotes mRNA export - NSUN2 as the methyltransferase and ALYREF as an m(5)C reader. Cell Res. 2017;27:606–25.
[14] Wang T, Kong S, Tao M, Ju S. The potential role of RNA N6-methyladenosine in Cancer progression. Mol Cancer. 2020;19:88.
[15] Dominissini D, Moshitch-Moshkovitz S, Schwartz S, et al. Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. Nature. 2012;485:201–6.
[16] Meyer KD, Saletore Y, Zumbo P, Elemento O, Mason CE, Jaffrey SR. Comprehensive analysis of mRNA methylation reveals enrichment in 3' UTRs and near stop codons. Cell. 2012;149:1635–46.
[17] Yang Y, Hsu PJ, Chen YS, Yang YG. Dynamic transcriptomic m(6)A decoration: writers, erasers, readers and functions in RNA metabolism. Cell Res. 2018;28:616–24.
[18] Zhang Z, Luo K, Zou Z, et al. Genetic analyses support the contribution of mRNA N(6)-methyladenosine (m(6)A) modification to human disease heritability. Nat Genet. 2020;52:939–49.
[19] Huang H, Weng H, Chen J. The biogenesis and precise control of RNA m(6)A methylation. Trends Genet. 2020;36:44–52.
[20] Ingolia NT, Lareau LF, Weissman JS. Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. Cell. 2011;147:789–802.
[21] Wang LC, Chen SH, Shen XL, et al. M6A RNA Methylation Regulator HNRNPC Contributes to Tumorigenesis and Predicts Prognosis in Glioblastoma Multiforme. Front Oncol. 2020;10:536875.
[22] Patil DP, Chen CK, Pickering BF, et al. m(6)A RNA methylation promotes XIST-mediated transcriptional repression. Nature. 2016;537:369–73.
[23] Widagdo J, Zhao QY, Kempen MJ, et al. Experience-dependent accumulation of N6-methyladenosine in the prefrontal cortex is associated with memory processes in mice. J Neurosci. 2016;36:6771–7.
[24] Clough E, Barrett T. The gene expression omnibus database. Methods Mol Biol. 2016;1418:93–110.
[25] Duke DC, Moran LB, Pearce RK, Graeber MB. The medial and lateral substantia nigra in Parkinson's disease: mRNA profiles associated with higher brain tissue vulnerability. Neurogenetics. 2007;8:83–94.
[26] Mutez E, Larvor L, Leprêtre F, et al. Transcriptional profile of Parkinson blood mononuclear cells with LRRK2 mutation. Neurobiol Aging. 2011;32:1839–48.
[27] Rigatti SJ. Random forest. J Insur Med. 2017;47:31–9.
[28] Huang S, Cai N, Pacheco PP, Narrandes S, Wang Y, Xu W. Applications of support vector machine (SVM) learning in cancer genomics. Cancer Genomics Proteomics. 2018;15:41–51.

[29] Wilkerson MD, Hayes DN. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. Bioinformatics. 2010;26:1572–3.
[30] Zhang B, Wu Q, Li B, Wang D, Wang L, Zhou YL. m(6)A regulator-mediated methylation modification patterns and tumor microenvironment infiltration characterization in gastric cancer. Mol Cancer. 2020;19:53.
[31] Tysnes OB, Storstein A. Epidemiology of Parkinson's disease. J Neural Transm (Vienna). 2017;124:901–5.
[32] Morris HR, Spillantini MG, Sue CM, Williams-Gray CH. The pathogenesis of Parkinson's disease. Lancet. 2024;403:293–304.
[33] Jiang X, Liu B, Nie Z, et al. The role of m6A modification in the biological functions and diseases. Signal Transduct Target Ther. 2021;6:74.
[34] Zhang N, Ding C, Zuo Y, Peng Y, Zuo L. N6-methyladenosine and neurological diseases. Mol Neurobiol. 2022;59:1925–37.
[35] Zhou J, Han Y, Hou R. Potential role of N6-methyladenosine modification in the development of Parkinson's disease. Front Cell Dev Biol. 2023;11:1321995.
[36] Qin L, Min S, Shu L, et al. Genetic analysis of N6-methyladenosine modification genes in Parkinson's disease. Neurobiol Aging. 2020;93:143.e9–13.
[37] Quan W, Li J, Liu L, et al. Influence of N6-Methyladenosine Modification Gene HNRNPC on Cell Phenotype in Parkinson's Disease. Parkinson's disease. 2021;2021:9919129.
[38] Liao J, Wei Y, Liang J, et al. Insight into the structure, physiological function, and role in cancer of m6A readers-YTH domain-containing proteins. Cell Death Discovery. 2022;8:137.
[39] Hou G, Zhao X, Li L, et al. SUMOylation of YTHDF2 promotes mRNA degradation and cancer progression by increasing its binding affinity with m6A-modified mRNAs. Nucleic Acids Res. 2021;49:2859–77.
[40] Wojtas MN, Pandey RR, Mendel M, Homolka D, Sachidanandam R, Pillai RS. Regulation of m(6)A Transcripts by the 3'→5' RNA Helicase YTHDC2 is essential for a successful meiotic program in the mammalian germline. Mol Cell. 2017;68:374–87.e12.
[41] Kretschmer J, Rao H, Hackert P, Sloan KE, Höbartner C, Bohnsack MT. The m(6)A reader protein YTHDC2 interacts with the small ribosomal subunit and the 5'-3' exoribonuclease XRN1. RNA. 2018;24:1339–50.
[42] Li J, Mahajan A, Tsai MD. Ankyrin repeat: a unique motif mediating protein-protein interactions. Biochemistry. 2006;45:15168–78.
[43] Liu N, Dai Q, Zheng G, He C, Parisien M, Pan T. N(6)-methyladenosine-dependent RNA structural switches regulate RNA-protein interactions. Nature. 2015;518:560–4.
[44] Wu Y, Zhao W, Liu Y, et al. Function of HNRNPC in breast cancer cells by controlling the dsRNA-induced interferon response. EMBO J. 2018;37:e99017.
[45] Zhang Y, Chen W, Pan T, Wang H, Zhang Y, Li C. LBX2-AS1 is activated by ZEB1 and promotes the development of esophageal squamous cell carcinoma by interacting with HNRNPC to enhance the stability of ZEB1 and ZEB2 mRNAs. Biochem Biophys Res Commun. 2019;511:566–72.
[46] Huang H, Han Y, Zhang C, et al. HNRNPC as a candidate biomarker for chemoresistance in gastric cancer. Tumour Biol. 2016;37:3527–34.
[47] Ruzzenente B, Metodiev MD, Wredenberg A, et al. LRPPRC is necessary for polyadenylation and coordination of translation of mitochondrial mRNAs. EMBO J. 2012;31:443–56.
[48] Oláhová M, Hardy SA, Hall J, et al. LRPPRC mutations cause early-onset multisystem mitochondrial disease outside of the French-Canadian population. Brain. 2015;138(Pt 12):3503–19.
[49] Jiang X, Li X, Huang H, et al. Elevated levels of mitochondrion-associated autophagy inhibitor LRPPRC are associated with poor prognosis in patients with prostate cancer. Cancer. 2014;120:1228–36.
[50] Tian T, Ikeda J, Wang Y, et al. Role of leucine-rich pentatricopeptide repeat motif-containing protein (LRPPRC) for anti-apoptosis and tumourigenesis in cancers. Eur J Cancer. 2012;48:2462–73.
[51] Gao W, Xua J, Wang F, et al. Mitochondrial proteomics approach reveals voltage-dependent anion channel 1 (VDAC1) as a potential biomarker of gastric cancer. Cell Physiol Biochem. 2015;37:2339–54.
[52] Fahrmann JF, Grapov D, Phinney BS, et al. Proteomic profiling of lung adenocarcinoma indicates heightened DNA repair, antioxidant mechanisms and identifies LASP1 as a potential negative predictor of survival. Clin Proteomics. 2016;13:31.
[53] Bennett JP, Jr., Keeney PM. Alzheimer's and Parkinson's brain tissues have reduced expression of genes for mtDNA OXPHOS Proteins, mitobiogenesis regulator PGC-1α protein and mtRNA stabilizing protein LRPPRC (LRP130). Mitochondrion. 2020;53:154–7.
[54] Amutha P, Rajkumar T. Role of insulin-like growth factor, insulin-like growth factor receptors, and insulin-like growth factor-binding proteins in ovarian cancer. Indian J Med Paediatr Oncol. 2017;38:198–206.
[55] Tong L, Cheng J, Zuo H, Li J. MicroRNA-197 promotes proliferation and inhibits apoptosis of gallbladder cancer cells by targeting insulin-like growth factor-binding protein 3. Adv Clin Exp Med. 2021;30:661–72.
[56] Li M, Wu W, Deng S, Shao Z, Jin X. TRAIP modulates the IGFBP3/AKT pathway to enhance the invasion and proliferation of osteosarcoma by promoting KANK1 degradation. Cell Death Dis. 2021;12:767.
[57] Watanabe K, Uemura K, Asada M, et al. The participation of insulin-like growth factor-binding protein 3 released by astrocytes in the pathology of Alzheimer's disease. Mol Brain. 2015;8:82.