Clin. Lab. 2026;72:XXX-XXX

ORIGINAL ARTICLE

Identification of Glycolysis-Related Diagnostic Biomarkers for Amyotrophic Lateral Sclerosis Using Machine Learning

Ping Ye [1, 2], Binhan Guo [1, 2]

[1] *Department of Laboratory Medicine, West China Second University Hospital, Sichuan University, Chengdu, China*
[2] *Key Laboratory of Birth Defects and Related Diseases of Women and Children (Sichuan University), Ministry of Education, Chengdu, China*

ABSTRACT

Background: **Glycometabolism has been implicated in the pathogenesis of amyotrophic lateral sclerosis (ALS), yet the precise molecular mechanisms underlying this association remain poorly understood. The identification of reliable biomarkers for ALS diagnosis represents a critical unmet need in clinical practice, as early detection and intervention could significantly improve patient outcomes.**
Methods: **We employed a comprehensive analytical approach combining two-sample Mendelian randomization analysis to investigate the causal relationship between blood glucose levels and ALS. Additionally, we integrated differential expression analysis, multiple machine learning algorithms, and correlation analyses to identify potential diagnostic biomarkers for ALS. The machine learning framework utilized gradient boosting tree methodology to construct predictive models, with performance evaluation conducted through cross-validation procedures.**
Results: **Mendelian randomization analysis demonstrated a significant negative causal relationship between blood glucose levels and ALS risk. Through bioinformatic analysis and machine learning approaches, we successfully identified candidate genes and constructed a high-performance predictive model using gradient boosting tree methodology, achieving an average area under the curve (AUC) of 0.8782 in cross-validation. Validation studies utilizing both bulk and single-cell RNA sequencing datasets revealed that *COL5A1* and *VCAN* genes play significant roles in ALS pathogenesis, likely through their involvement in glycolytic pathways.**
Conclusions: **Our findings provide novel insights into the molecular mechanisms linking glycometabolism and ALS, while identifying potential diagnostic biomarkers for the disease. The identified genes, *COL5A1* and *VCAN*, represent promising targets for further investigation in ALS pathogenesis. However, the clinical translation of these findings requires validation through additional datasets and prospective clinical trials to establish their diagnostic utility and therapeutic potential.**
(Clin. Lab. 2026;72:xx-xx. DOI: 10.7754/Clin.Lab.2025.250766)

Correspondence:
Binhan Guo
Department of Laboratory Medicine
West China Second University Hospital
Sichuan University
Chengdu, 610041
China
Email: 2183983697@qq.com
ORCID: 0000-0002-1448-8022

Manuscript accepted August 14, 2025

KEYWORDS

amyotrophic lateral sclerosis, glycolysis, machine learning, identification

INTRODUCTION

Amyotrophic lateral sclerosis (ALS) is a chronic, progressive, degenerative disease involving the upper and lower motor neurons, including the innervated muscles in the trunk, limbs, head, and face [1,2]. It causes progressive muscle weakness, atrophy, and bundle tremors caused by combined damage to the upper and lower motor neurons, leading to death, typically within 2 -

5 years. In recent years, genetic variations associated with ALS development have been identified, in genes including *SOD1*, *TARDBP*, *FUS*, and *C9orf72*; these are being increasingly used as biomarkers for disease diagnosis [3-5]. The pathophysiological process of ALS is multifactorial, with intricate connections among genetics, inflammation, immunity, and metabolism. Novel biomarkers are required to achieve early diagnosis and treatment and to improve patient prognosis.

Glucose is the main energy source used by the human brain to fuel various cellular metabolic activities and nervous system functions, and the brain is the main consumer of blood sugar, especially during rest [6]. Abnormal glucose metabolism can lead to the accumulation of glycogen in tissues, which can have detrimental effects. A study in mice showed that glycogen accumulation in reactive astrocytes contributes to neurotoxicity and disease progression in ALS [7]. Another study revealed substantial differences in glucose metabolism in the lateral peridigital area, thalamus, and brainstem of patients with ALS harboring *C9orf72* mutations compared to healthy controls [8]. In addition, relevant studies have suggested that the decrease in glucose transport-related protein levels observed in patients with ALS affects glucose transport and glycolysis [9,10]. The findings of these studies indicate an abnormal glycometabolic status in patients with ALS; however, the results of observational studies can be affected by unmeasured confounding factors. Therefore, we used Mendelian randomization (MR) [11,12] to investigate the association between blood glucose level and ALS development and subsequently analyzed glycolysis-related gene expression in patients with ALS compared with that in healthy controls, with the aim to identify valuable biomarkers of this disease.

Recently, bioinformatics has been extensively utilized in multiple research fields to identify biomarkers, screen drug targets, and interpret biological functions. This technique involves the use of massive amounts of raw data, including microarray, sequencing, and mass spectrometry data, mined from public databases [13]. However, analysis from a single data source may lead to false-positive results, making it challenging to obtain reliable conclusions. Machine learning, a computer science method employed to identify complex patterns in data, is widely used to predict or classify new unknown data based on learning from existing data [14]. The process of machine learning, which uses a combination of different datasets, can substantially reduce the false-positive rate compared with the results obtained from a single dataset, which helps to better understand human health and disease.

Therefore, in the present study, we screened candidate genes and constructed a diagnostic model based on machine learning by jointly analyzing multiple datasets from public sources using bioinformatics to investigate the potential mechanisms underlying the association between glycolysis and ALS development. The aim of the study, to identify novel biomarkers of ALS, was achieved, although further verification of our findings will be required.

MATERIALS AND METHODS

MR analysis

To explore the causality between blood sugar levels and ALS, genetic data on blood sugar level (bbj-a-10; sample size, 93,146; total number of single nucleotide polymorphisms, 6,108,953) and ALS (ieu-a-1085; sample size, 36,052; total number of single nucleotide polymorphisms, 7,740,345) were obtained from the Integrative Epidemiology Unit Open genome-wide association study database (https://gwas.mrcieu.ac.uk/). Blood sugar was used as the exposure factor, and ALS was the outcome variable. Two-sample MR analysis was performed using inverse-variance weighted, weighted median, MR-Egger, simple mode, and weighted mode approaches. If more than three of these approaches yielded similar results (using $p < 0.05$ as the significance threshold), the results of the two-sample MR were considered reliable.

Data source and screening of intersected genes

The dataset GSE209696, which is based on the GPL24676 platform (Illumina NovaSeq 6000), includes high-throughput sequencing data from 239 patients with ALS and 29 healthy controls. The dataset GSE112681 contains RNA sequencing data obtained from the peripheral blood of 397 patients with ALS and 645 healthy controls, based on the GPL6947 (Illumina HumanHT-12 V3.0 expression beadchip) and GPL10558 (Illumina HumanHT-12 V4.0 expression beadchip) platforms. The dataset GSE234297 contains RNA sequencing data obtained from the peripheral blood of 96 patients with ALS and 48 healthy controls, based on the GPL11154 platform (Illumina HiSeq 2000).

The dataset GSE233881 contains chip data obtained from the peripheral blood of 96 patients with ALS and 48 healthy controls, based on the GPL64980 platform (Agilent-014850 Whole Human Genome Microarray 4×44K G4112F). The dataset GSE264012 contains single-cell RNA sequencing (scRNA) data obtained from day 90 induced pluripotent stem cell-derived cerebral organoid cells from three patients with ALS and three healthy controls, based on the GPL30173 platform (NextSeq 2000). All datasets mentioned above were freely downloaded from the Gene Expression Omnibus database (http://www.ncbi.nlm.nih.gov/geo). The R software (The R Foundation for Statistical Computing, Vienna, Austria) package “Limma” was utilized to screen for differentially expressed genes (DEGs) in GSE209696, with thresholds set as fold change > 1 and $p < 0.05$. Glycolysis genes (GGs) obtained from the hallmark gene set of the Human Molecular Signatures Database (https://www.gsea-msigdb.org) were assessed for intersection with differentially expressed genes (DEGs) from GSE209696. The R packages “Seurat,”

"SingleR," and "Harmony" were used to analyze the GSE264012 dataset.

Construction of machine learning models

The genes obtained from the intersection of differentially expressed genes (DEGs) from GSE209696 and glycolysis genes (GGs) were used as the independent variable, while the group labels of samples (from patients with ALS and healthy controls) were used as the dependent variable. Data were entered into the logistic (glmnet) [32], multilayer perceptron neural network (mlp) [33], and eXtreme Gradient boosting tree (XGBoost) [32] models constructed using the R package "tidymodels" [34]. First, 1,042 individuals from GSE112681 were randomly divided into training and testing sets for three-fold outer cross-validation (CV). Second, each fold of the training sets for outer cross-validation (CV) was divided into training and testing sets for 10-fold inner cross-validation (CV). Third, the hyperparameters of the model were optimized by inner cross-validation (CV) of the training sets. Subsequently, the best hyperparameter identified by inner cross-validation (CV) was applied to fit each fold of the training sets for outer cross-validation (CV) and was finally evaluated using the testing sets for outer cross-validation (CV). Predictive performance was evaluated in each fold using different statistical parameters, including accuracy, balanced accuracy, precision, recall, F1 (harmonic mean of precision and recall), area under the receiver operating characteristic curve, and area under the precision versus recall curve.

Candidate gene screening and classification model construction

To further screen independent variables with a high contribution to the models in machine learning, additional datasets were utilized to explore the intersected differentially expressed genes (DEGs) from GSE209696 and glycolysis genes (GGs). The expression of these intersected genes in the dataset GSE234297 was explored and visualized using a box plot to identify genes that were significantly differentially expressed between the ALS and healthy control groups. In addition, Pearson correlation analysis was performed on the same intersected genes in the dataset GSE233881 to find genes with a strong correlation (Pearson correlation coefficient, $0.6 < |R| < 0.8$). Thereafter, genes screened from GSE234297 and GSE233881 were combined as candidate genes. These candidate genes were entered into the eXtreme Gradient boosting tree (XGBoost) model as independent variables, to construct a model for the classification of ALS.

Expression of candidate genes in the single-cell RNA sequencing (scRNA) dataset

The single-cell RNA sequencing (scRNA) dataset was processed by filtering out low-quality cells with less than 200 genes or mitochondrial genes exceeding 25%. Next, we identified 2,000 highly variable genes and performed dimensionality reduction and clustering on the cell clusters. According to the results of the elbow plot, the first 50 dimensions were selected for follow-up analysis based on suitable principal component (PC) values. Uniform manifold approximation and projection (UMAP) was utilized to identify key structures in a high-dimensional space on the identified cell clusters and visualize them in a low-dimensional space. The cell clusters were annotated according to the highly expressed genes in each cluster and previously reported cell markers. Finally, the candidate genes for construction of the final model were visualized in different cell clusters using a dot plot.

Statistic

All statistical analyses were performed using R software 4.4.1. Student's *t*-test or Wilcoxon were utilized to investigate the difference between two groups. The parameter of correlation analysis between the variables was set as Pearson. All statistical p-values were two-sided, and $p < 0.05$ or one asterisk (*) was regarded as statistical significance.

RESULTS

Mendelian randomization (MR) analysis

The Mendelian randomization (MR) analysis showed that increased blood sugar level was associated with a reduction in the risk of ALS using the weighted median (odds ratio [OR]: 0.696, 95% confidence interval [CI]: 0.661 - 0.732, $p < 0.001$), inverse-variance weighted (OR: 0.686, 95% CI: 0.642 - 0.733, $p < 0.001$), simple mode (OR: 0.683, 95% CI: 0.641 - 0.728, $p < 0.001$), and weighted mode (OR: 0.698, 95% CI: 0.662 - 0.735, $p < 0.001$) approaches. Only the MR-Egger approach yielded an insignificant result (OR: 0.926, 95% CI: 0.745 - 1.150, $p = 0.488$), as presented in Figure 1.

Intersected gene screening and machine learning

A total of 3,171 differentially expressed genes (DEGs) (1,103 upregulated and 2,068 downregulated in patients with ALS) were identified using the GSE209696 dataset, as shown in Figure 2A. Out of these differentially expressed genes (DEGs), 31 were glycolysis genes (GGs); the expression levels of these genes are displayed as a heat map in Figure 2B. These 31 genes were used as independent variables for machine learning, and the dataset GSE112681 was randomly categorized into training and testing sets for three-fold cross-validation (CV). Figures 3A - C show the results of each fold for the classification of ALS using the logistic (glmnet), eXtreme Gradient boosting tree (XGBoost), and multilayer perceptron neural network (mlp) models. Supplementary Tables S1 - 3 online present the average statistical parameters used to evaluate the predictive performance of the different models for cross-validation (CV) of each fold. The average cross-validation (CV) area under the curve (AUC) value was 0.6922 for the glmnet

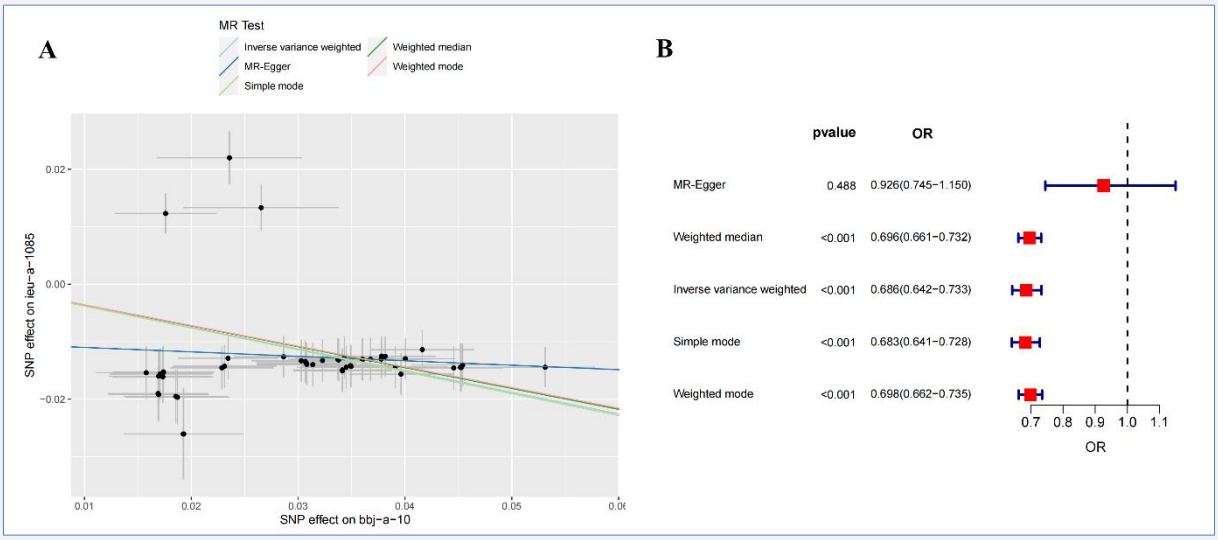

Figure 1. Association between blood sugar level and risk of amyotrophic lateral sclerosis.

a: Scatter plot of MR analyses performed using MR-Egger, weighted median, inverse-variance weighted, simple mode, and weighted mode approaches. b: Forest plot of statistical results. MR: Mendelian randomization, OR: odds ratio, SNP: single nucleotide polymorphism.

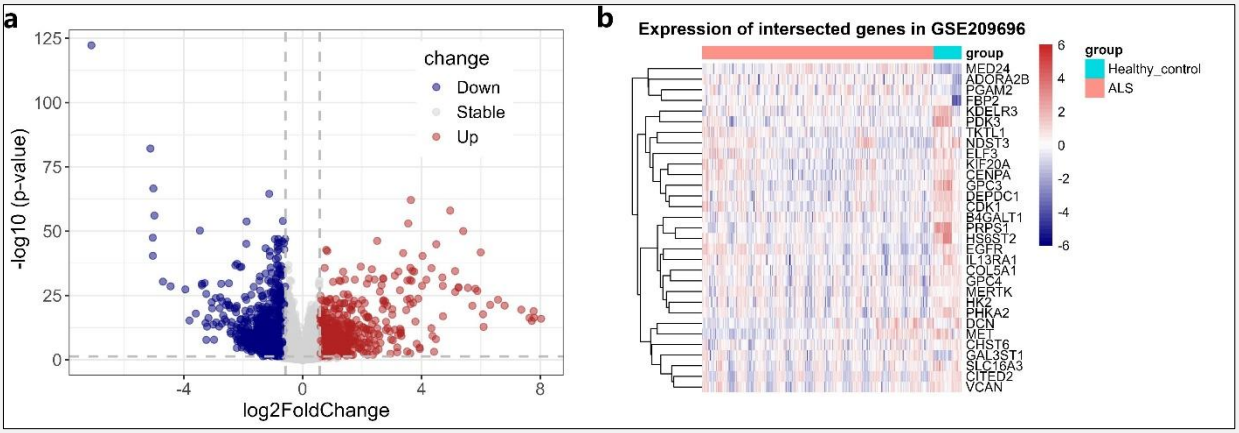

Figure 2. Genes differentially expressed between patients with ALS and healthy controls.

a: Volcano plot of the 3,171 differentially expressed genes identified from the dataset GSE209696; 1,103 genes were upregulated and 2,068 genes were downregulated in patients with ALS. b: Heat map of 31 differentially expressed genes that are also glycolysis genes. Red indicates genes whose expression was upregulated in the ALS group compared with those in the healthy control group; blue indicates genes whose expression was downregulated in the ALS group compared with those in the healthy control group. ALS: amyotrophic lateral sclerosis.

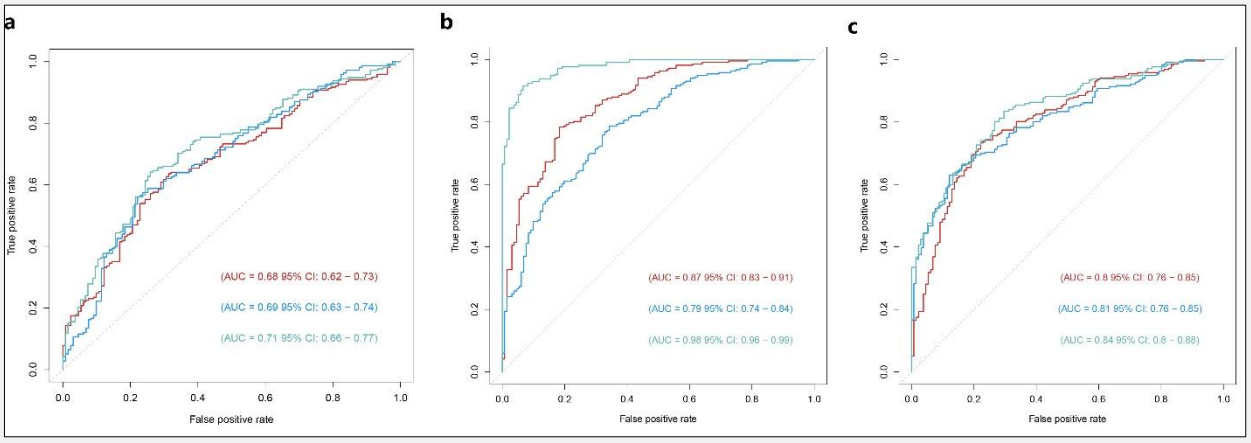

Figure 3. Ability of various models to predict amyotrophic lateral sclerosis classification.

Receiver operating characteristic curves showing the predictive ability of the a: logistic model (glmnet), b: eXtreme Gradient boosting tree (XGBoost) model, and c: multilayer perceptron neural network model (mlp) in the testing set. The different line colors represent the different folds used for cross-validation. AUC: area under the curve, CI: confidence interval.

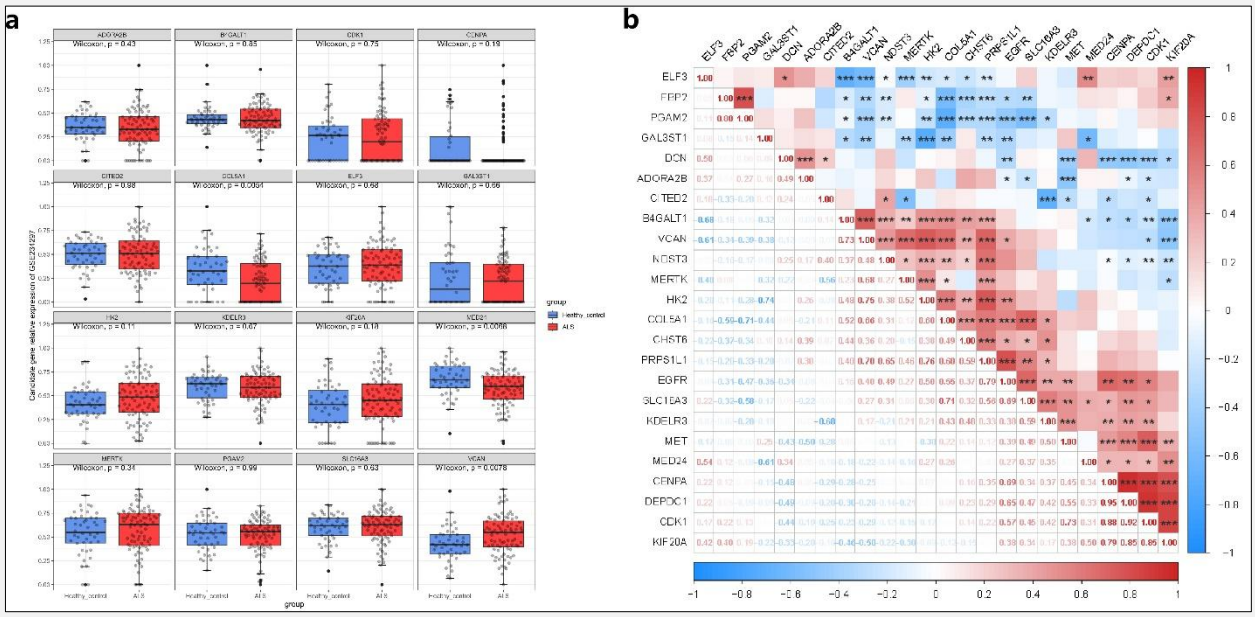

Figure 4. Candidate gene expression.

a: Expression of some of the "model-genes" in the ALS and healthy control groups from the dataset GSE234297. Gene expression levels in the box plot were standardized through 0 - 1 range. Gene expression levels were compared between the groups using Wilcoxon's test, with $p < 0.05$ considered to indicate statistical significance. b: Correlation of "model-gene" expression in the dataset GSE233881. Red indicates a positive correlation; blue indicates a negative correlation. *** $p < 0.001$, ** $p < 0.01$, and * $p < 0.05$. ALS: amyotrophic lateral sclerosis.

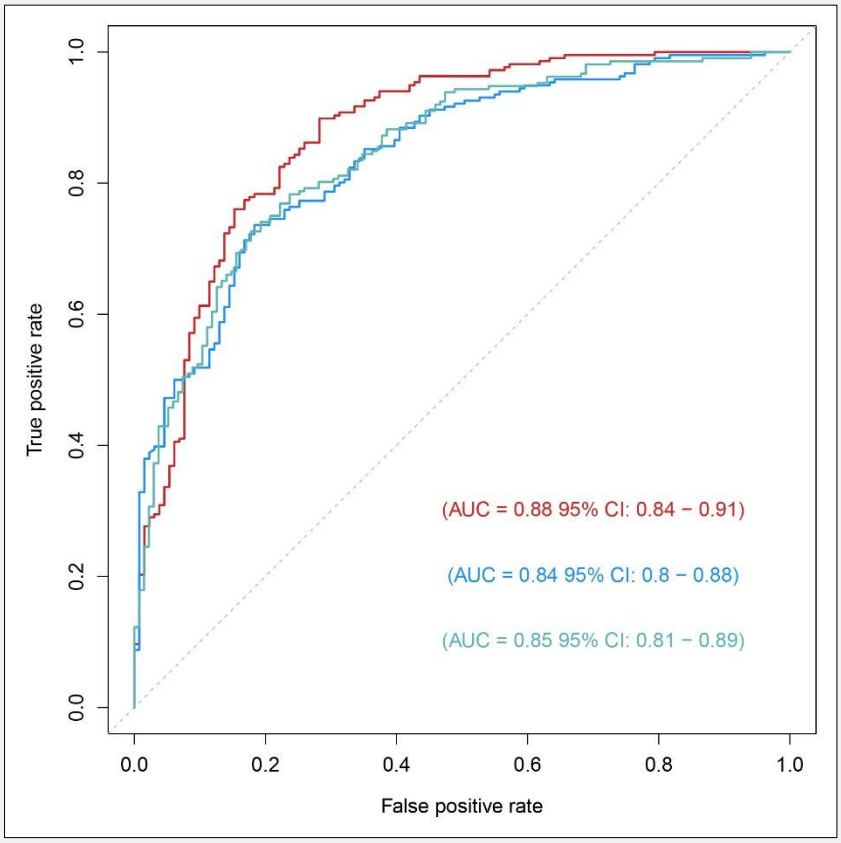

Figure 5. Ability of the eXtreme Gradient boosting tree (XGBoost) model constructed using candidate genes to predict amyotrophic lateral sclerosis.

Predictive ability was evaluated by plotting receiver operating characteristic curves for the testing set. The different line colors represent the different folds used for cross-validation. AUC: area under the curve, CI: confidence interval.

model, 0.8782 for the eXtreme Gradient boosting tree (XGBoost) model, and 0.8161 for the mlp model. The average accuracy of cross-validation (CV) area under the curve (AUC) value was 0.6430 for the glmnet model, 0.8071 for the eXtreme Gradient boosting tree (XGBoost) model, and 0.7303 for the mlp model.

Candidate gene screening and classification model construction

Based on the results obtained from machine learning, 24 genes (*ADORA2B*, *B4GALT1*, *CDK1*, *CENPA*, *CHST6*, *CITED2*, *COL5A1*, *DCN*, *DEPDC1*, *EGFR*, *ELF3*, *FBP2*, *GAL3ST1*, *HK2*, *KDELR3*, *KIF20A*, *MED24*, *MERTK*, *MET*, *NDST3*, *PGAM2*, *PRPS1L1*, *SLC16A3*, and *VCAN*) were screened after data pre-processing before model construction, called "model-genes." The expression of some of these "model-genes" in the dataset GSE234297 is shown in Figure 4A in the form of a box plot, which shows significant differences in the expression of genes such as *COL5A1, VCAN,* and *MED24* between the ALS and healthy control groups. In addition, the correlation analysis of the expression of "model-genes" in the dataset GSE233881 showed a negative correlation between *COL5A1* and *PGAM2* ($R = -0.71$, $p < 0.001$) and positive correlations between *COL5A1* and *SLC16A3* ($R = 0.71$, $p < 0.001$), *VCAN* and *B4GALT1* ($R = 0.73$, $p < 0.001$), *VCAN* and *MERTK* ($R = 0.68$, $p < 0.001$), *VCAN* and *HK2* ($R = 0.75$, $p < 0.001$), *COL5A1* and *VCAN* ($R = 0.66$, $p < 0.001$), and *VCAN* and *PRPS1L1* ($R = 0.70$, $p < 0.001$), as presented in Figure 4B. Based on these results, nine candidate genes (*COL5A1*, *PGAM2*, *VCAN*, *SLC16A3*, *B4GALT1*, *MERTK*, *HK2*, *PRPS1L1*, and *MED24*) were selected as independent variables to construct the eXtreme Gradient boosting tree (XGBoost) model using the dataset GSE-112681. Figure 5 shows the result of each fold for the classification of ALS using the eXtreme Gradient boosting tree (XGBoost) model built using candidate genes,

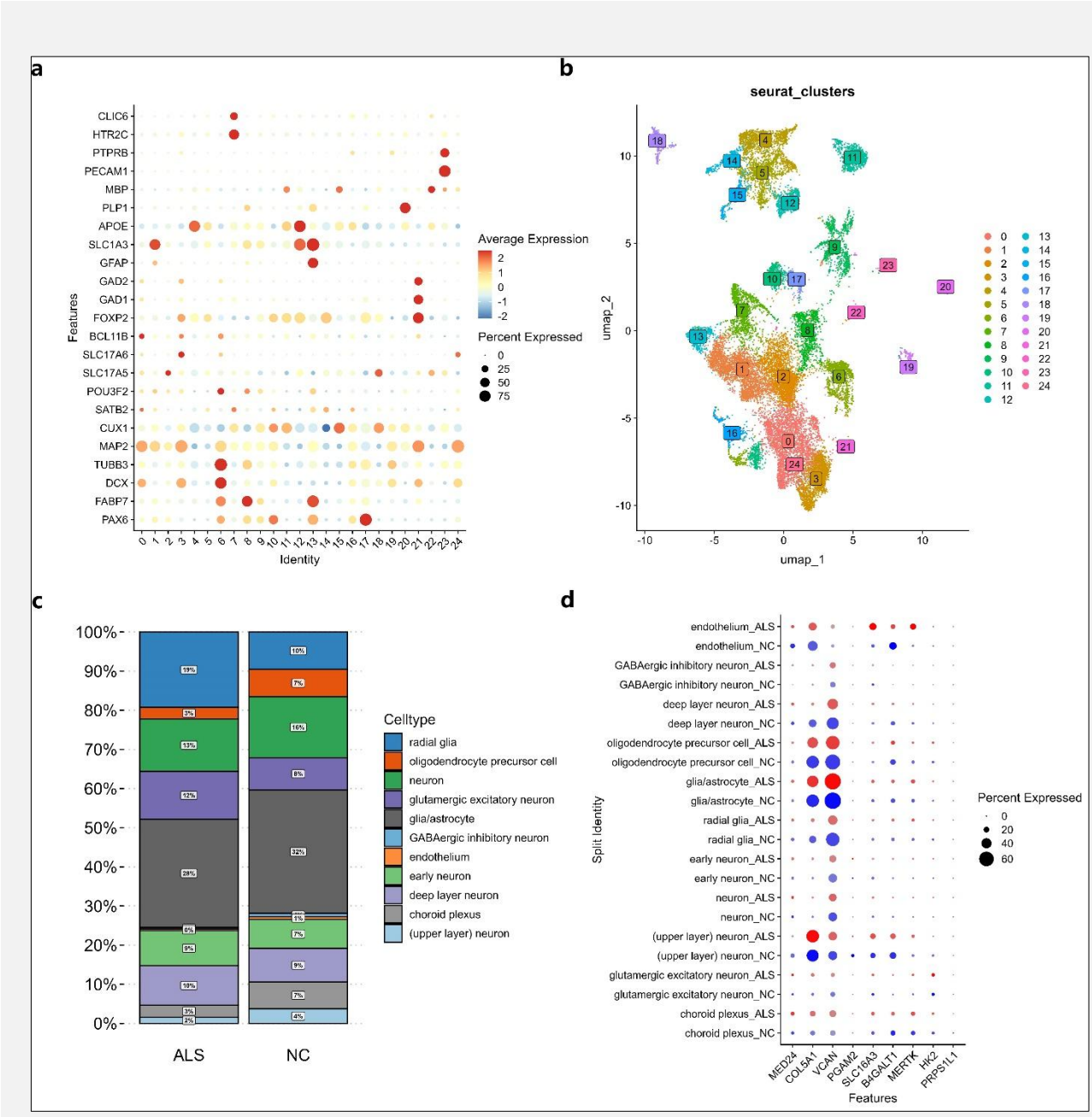

Figure 6. Expression of candidate genes in the single-cell RNA sequencing dataset.

a: Expression of canonical markers used for cell annotation in each cluster. b: UMAP clustering through each cluster. c: Cell type proportions in NC and ALS groups. d: Distribution of candidate genes in 11 cell clusters. Red represents the ALS group; blue represents the normal control group. The size of the dot indicates the level of gene expression. ALS: amyotrophic lateral sclerosis, NC: normal control.

and the average statistical parameters used to evaluate its predictive performance are presented in Supplementary Table S4 online.

Expression of candidate genes in the single-cell RNA sequencing dataset

Using the single-cell RNA sequencing (scRNA) dataset, a total of 22,912 cells from six samples (10,955 cells from patients with ALS and 11,957 cells from healthy controls) passed quality control and were used for fur-

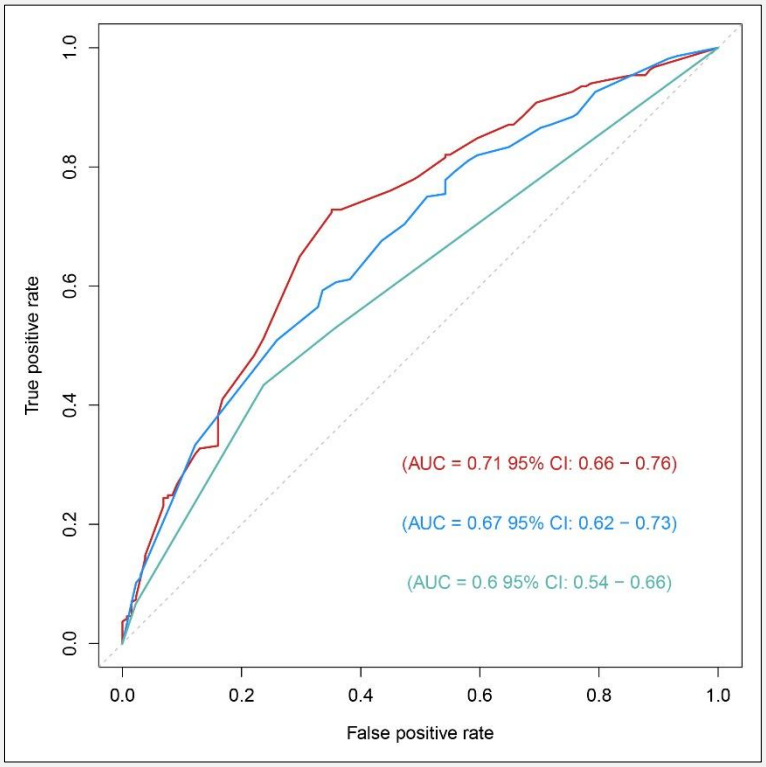

Figure 7. Ability of the XGBoost model constructed using *COL5A1* and *VCAN* to predict amyotrophic lateral sclerosis.

Predictive ability was evaluated by plotting receiver operating characteristic curves for the testing set. The different line colors represent the different folds used for cross-validation. AUC: area under the curve, CI: confidence interval.

ther analysis. After removing batch effects and normalizing the data, 25 cell clusters were obtained at a resolution of 0.7. Combining clusters resulted in a total of 11 cell clusters: radial glia, early neurons, neurons, (upper layer) neurons, glutamatergic excitatory neurons, deep-layer neurons, GABAergic inhibitory neurons, glia/astrocytes, oligodendrocyte precursor cells, endothelium, and choroid plexus (Figure 6A). The marker genes for each cell cluster are listed in Supplementary Table S5 online. Radial glia accounted for 19% of the total cells in the ALS group, significantly higher than the proportion of these cells in the healthy control group (10%) (Figure 6B). We then visualized the expression of *COL5A1*, *PGAM2*, *VCAN*, *SLC16A3*, *B4GALT1*, *MERTK*, *HK2*, *PRPS1L1*, and *MED24* in the ALS and healthy control groups in the 11 cell clusters. The differential expression of *COL5A1* was mainly distributed in deep layer neurons and radial glia, and the differential expression of *VCAN* was mainly in radial glia (Figure 6C).

Construction of classification model using *COL5A1* and *VCAN*

Following validation of the machine learning results using bulk RNA/scRNA data, two genes (*COL5A1* and *VCAN*) were selected as independent variables for the construction of the eXtreme Gradient boosting tree (XGBoost) model using the dataset GSE112681. Figure 7 shows the results of each fold for ALS classification using this model; the average statistical parameters used to evaluate the predictive performance are shown in Supplementary Table S6 online.

DISCUSSION

The results of our Mendelian randomization (MR) analysis provided evidence that each unit increase in genetically predicted natural-log transformed blood sugar level conferred a reduced risk of ALS, which was consistent and robust across four sensitivity analyses (weighted median, inverse-variance weighted, simple, and weighted modes). Similarly, previous studies have

shown that the incidence of ALS is lower in patients with type 2 diabetes and obesity than in the general population, especially in men and patients aged 60 years or above [15]. However, a different study indicated that patients with ALS who carry *C9ORF72* and *SOD1* gene mutations exhibit significant differences in blood glucose levels and body mass index, which reflect different cardiovascular risks [16]. This finding suggests that the normal upregulation of glycolysis in response to an increase in blood sugar levels may be disrupted in patients with ALS. Our Mendelian randomization (MR) analysis indicated a potential relationship between blood sugar levels and the risk of ALS to a certain extent, which laid the foundation for further analysis of the association between glycolysis and ALS.

Bioinformatics identified 31 genes at the intersection between differentially expressed genes (DEGs) from GSE209696 and glycolysis genes (GGs), indicating a possible complex interaction mechanism in ALS. Recently, the application of machine learning in medicine has provided substantial information by mining potential data from complex medical cases and conducting statistical analyses using different algorithms [17,18]. Considering that the 31 genes identified in this study are differentially expressed in ALS and related to glycolysis, a model was developed based on another dataset and using three different machine learning algorithms through cross-validation (CV), eliminating possible errors caused by the manual selection of the same datasets. In addition, genes were screened for inclusion in the model by analyzing the correlation between independent variables and the characteristics of zero variance during data preparation. Our machine learning results indicated that the model constructed based on the gradient boosting tree had the best classification performance for ALS, followed by the multilayer perceptron neural network model; the logistics classification model exhibited the poorest performance. Similarly, different datasets correspond to applicable machine learning algorithms, as reflected in many recent studies [19-21]. Our study revealed genes differentially expressed in ALS and correlations between these genes. Although no direct research has suggested a relationship among *COL5A1, VCAN*, and ALS, the following studies have provided indirect evidence. One study identified *VCAN* as a candidate gene for constructing a model of Alzheimer's disease (AD), another neurodegenerative disease [22], suggesting a potential role for *VCAN* in ALS. Similarly, Rahman et al. [23] identified *PGAM2* as a molecular biomarker that helped predict the progression of ischemic stroke and AD. Furthermore, Tang et al. [24] reported that upregulation of *B4GALT1* expression may contribute to AD pathogenesis through its involvement in complex N-linked glycan formation and galactosylation. Although associations between these genes and ALS development have not been reported, AD and ALS may have similar pathogeneses.

Research suggests that long-term excessive intake of sugar may affect metabolism, leading to the generation and accumulation of inflammatory mediators, thereby causing mild chronic inflammation [25-27]. Chronic neuroinflammation is often observed in patients with neurodegenerative diseases. In addition, it has been suggested that the tumor-associated macrophage-driven pathway contributes to neurodegenerative diseases by controlling inflammatory responses through anti-inflammatory actions [28,29], which indirectly indicates a potential association between *MERTK* and ALS development. Moreover, glucose metabolism reprogramming in microglia, which triggers chronic inflammation upon stimulation, may interact with the immune system, ultimately leading to disease [30,31]. Consistent with this finding, the abnormal expression of *COL5A1* and *VCAN* in the radial glia revealed by the present study suggests that they may affect the pathogenesis of ALS by participating in glycolysis. In addition, we found that *COL5A1* and *VCAN*, as “model-genes” involved in glycolysis, were differentially expressed between the ALS and healthy control groups in the dataset GSE234297, indicating their value as diagnostic markers. However, the eXtreme Gradient boosting tree (XGBoost) model based on *COL5A1* and *VCAN* has not yet demonstrated excellent performance in predicting ALS, possibly because of interactions with other related genes that require further in-depth exploration.

In conclusion, the gradient boosting tree machine learning model indicated the diagnostic value of nine candidate genes in patients with ALS. However, this study has some limitations. First, some errors may have occurred in the interpretation of the construction of the model owing to the limited access of clinical information in public databases. In addition, more datasets and prospective clinical trials are required to validate the diagnostic value of the candidate genes. Finally, the use of dozens or even hundreds of different algorithms for the parallel modeling of data may lead to the identification of more suitable biomarkers for the clinical diagnosis of ALS.

Acknowledgment:
The authors would like to thank Editage (www.editage.cn) for English language editing.

Ethical Approval Statement:
Integrative Epidemiology Unit Open genome-wide association study database and GEO are public databases. Ethical approval was obtained from the patients involved in these databases. Users can download relevant data for free for research and publish relevant articles. Our study was based on open-source data, and therefore, there are no ethical issues or other conflicts of interest.

Data Availability Statement:
GSE209696 can be downloaded from https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE209696, GSE112681 from https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE112681, GSE234297 from https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE234297, GSE233881 from https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE233881, and GSE264012 from https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE264012.

Source of Funds:
The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Declaration of Interest:
The authors declare that they have no competing interests.

References:

1. Grad LI, Rouleau GA, Ravits J, Cashman NR. Clinical spectrum of amyotrophic lateral sclerosis (ALS). Cold Spring Harb Perspect Med 2017 Aug 1;7(8):a024117. (PMID: 28003278)

2. Feldman EL, Goutman SA, Petri S, et al. Amyotrophic lateral sclerosis. Lancet 2022 Oct 15;400(10360):1363-80. (PMID: 36116464)

3. Witzel S, Mayer K, Oeckl P. Biomarkers for amyotrophic lateral sclerosis. Curr Opin Neurol 2022 Oct 1;35(5):699-704. (PMID: 35942674)

4. Martinelli I, Ghezzi A, Zucchi E, et al. Predictors for progression in amyotrophic lateral sclerosis associated to SOD1 mutation: insight from two population-based registries. J Neurol 2023 Dec; 270(12):6081-92. (PMID: 37668704)

5. Sung W, Kim JA, Kim YS, et al. An analysis of variants in TARDBP in the Korean population with amyotrophic lateral sclerosis: comparison with previous data. Sci Rep 2023 Nov 1;13(1): 18805. (PMID: 37914747)

6. Mergenthaler P, Lindauer U, Dienel GA, Meisel A. Sugar for the brain: the role of glucose in physiological and pathological brain function. Trends Neurosci 2013 Oct;36(10):587-97. (PMID: 23968694)

7. Brewer MK, Torres P, Ayala V, et al. Glycogen accumulation modulates life span in a mouse model of amyotrophic lateral sclerosis. J Neurochem 2024 May;168(5):744-59. (PMID: 37401737)

8. De Vocht J, Van Weehaeghe D, Ombelet F, et al. Differences in cerebral glucose metabolism in ALS patients with and without C9orf72 and SOD1 mutations. Cells 2023 Mar 18;12(6):933. (PMID: 36980274)

9. Guo Z, Kindy MS, Kruman I, Mattson MP. ALS-linked Cu/Zn-SOD mutation impairs cerebral synaptic glucose and glutamate transport and exacerbates ischemic brain injury. J Cereb Blood Flow Metab 2000 Mar;20(3):463-8. (PMID: 10724110)

10. Manzo E, Lorenzini I, Barrameda D, et al. Glycolysis upregulation is neuroprotective as a compensatory mechanism in ALS. Elife 2019 Jun 10:8:e45114. (PMID: 31180318)

11. Davey Smith G, Hemani G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. Hum Mol Genet 2014 Sep 15;23(R1):R89-98. (PMID: 25064373)

12. Sekula P, Del Greco MF, Pattaro C, Köttgen A. Mendelian randomization as an approach to assess causality using observational data. J Am Soc Nephrol 2016 Nov;27(11):3253-65. (PMID: 27486138)

13. Gauthier J, Vincent AT, Charette SJ, Derome N. A brief history of bioinformatics. Brief Bioinform 2019 Nov 27;20(6):1981-96. (PMID: 30084940)

14. MacEachern SJ, Forkert ND. Machine learning for precision medicine. Genome 2021 Apr;64(4):416-25. (PMID: 33091314)

15. Skajaa N, Riahi EB, Szépligeti SK, et al. Type 2 diabetes, obesity, and risk of amyotrophic lateral sclerosis: A population-based cohort study. Brain Behav 2023 Jun;13(6):e3007. (PMID: 37073502)

16. Xia K, Witzel S, Witzel C, Klose V, et al. Mutation-specific metabolic profiles in presymptomatic amyotrophic lateral sclerosis. Eur J Neurol 2023 Jan;30(1):87-95. (PMID: 36169607)

17. Brinati D, Campagner A, Ferrari D, Locatelli M, Banfi G, Cabitza F. Detection of COVID-19 infection from routine blood exams with machine learning: a feasibility study. J Med Syst 2020 Jul 1; 44(8):135. (PMID: 32607737)

18. Mitterecker A, Hofmann A, Trentino KM, et al. Machine learning-based prediction of transfusion. Transfusion 2020 Sep;60(9): 1977-86. (PMID: 32596877)

19. Chen DL, Cai JH, Wang CCN. Identification of key prognostic genes of triple negative breast cancer by LASSO-based machine learning and bioinformatics analysis. Genes (Basel) 2022 May 18;13(5):902. (PMID: 35627287)

20. Liu Y, Jiang H, Kang T, et al. Platelets-related signature based diagnostic model in rheumatoid arthritis using WGCNA and machine learning. Front Immunol 2023 Jun 23;14:1204652. (PMID: 37426641)

21. Nazari E, Khalili-Tanha G, Asadnia A, et al. Bioinformatics analysis and machine learning approach applied to the identification of novel key genes involved in non-alcoholic fatty liver disease. Sci Rep 2023 Nov 22;13(1):20489. (PMID: 37993474)

22. Lin J, Chen J, Huang C. Systematic identification of key basement membrane related genes as potential new biomarkers in Alzheimer's disease. Clin Neurol Neurosurg 2024 Jan;236: 108094. (PMID: 38154381)

23. Rahman MR, Islam T, Shahjaman M, et al. Discovering biomarkers and pathways shared by Alzheimer's disease and ischemic stroke to identify novel therapeutic targets. Medicina (Kaunas) 2019 May 22;55(5):191. (PMID: 31121943)

24. Tang X, Tena J, Di Lucente J, et al. Transcriptomic and glycomic analyses highlight pathway-specific glycosylation alterations unique to Alzheimer's disease. Sci Rep 2023 May 15;13(1):7816. (PMID: 37188790)

25. Della Corte KW, Perrar I, Penczynski KJ, Schwingshackl L, Herder C, Buyken AE. Effect of dietary sugar intake on biomarkers of subclinical inflammation: A systematic review and meta-analysis of intervention studies. Nutrients 2018 May 12;10(5):606. (PMID: 29757229)

26. Freeman CR, Zehra A, Ramirez V, Wiers CE, Volkow ND, Wang GJ. Impact of sugar on the body, brain, and behavior. Front Biosci (Landmark Ed) 2018 Jun 1;23(12):2255-66. (PMID: 29772560)

27. Ma X, Nan F, Liang H, et al. Excessive intake of sugar: an accomplice of inflammation. Front Immunol 2022 Aug 31;13: 988481. (PMID: 36119103)

28. Vago JP, Amaral FA, van de Loo FAJ. Resolving inflammation by TAM receptor activation. Pharmacol Ther 2021 Nov;227: 107893. (PMID: 33992683)

29. Burstyn-Cohen T, Fresia R. TAM receptors in phagocytosis: beyond the mere internalization of particles. Immunol Rev 2023 Oct;319(1):7-26. (PMID: 37596991)

30. Gimeno-Bayón J, López-López A, Rodríguez MJ, Mahy N. Glucose pathways adaptation supports acquisition of activated microglia phenotype. J Neurosci Res 2014 Jun;92(6):723-31. (PMID: 24510633)

31. Huang Q, Wang Y, Chen S, Liang F. Glycometabolic reprogramming of microglia in neurodegenerative diseases: insights from neuroinflammation. Aging Dis 2024 May 7;15(3):1155-75. (PMID: 37611905)

32. Doupe P, Faghmous J, Basu S. Machine Learning for health services researchers. Value Health 2019 Jul;22(7):808-15. (PMID: 31277828)

33. Kuhn M, Johnson K. (2013) Applied Predictive Modeling. New York: Springer. http://dx.doi.org/10.1007/978-1-4614-6849-3

34. Kuhn M, Wickham H. Tidymodels: a Collection of Packages for Modeling and Machine Learning Using Tidyverse Principles. https://www.tidymodels.org

Additional material can be found online at:
http://supplementary.clin-lab-publications.com/250766/