Open Access |Research Article

Identification of complement-related genes in lung tissue after brain death and cardiac death based on machine learning

Corresponding Author: Jiangfa Li

Tel: +86-07732824373; Email: 247546160@qq.com

Liping Lei[1,2#]; Jiangfa Li[3*#]

[1]Department of Geriatric Medicine, The First Affiliated Hospital of Guilin Medical University, China.

[2]Department of Respiratory and Critical Care Medicine, Second Affiliated Hospital of Guilin Medical University, China.

[3]Division of Hepatobiliary Surgery, The First Affiliated Hospital of Guilin Medical University, Guilin, Guangxi 541001, China.

[#]These authors have been equally contributed to this article.

Abstract

Background and aims: The effects of brain death and cardiac death on lung tissue were different. In the pathological process of brain death and heart death, the activation and regulation of complement system show obvious differences. Thus, our study aims to explore the difference of complement-related genes between the two.

Methods: Using data from the Gene Expression Omnibus (GEO), we explored the expression profiles of Complement-Related Genes (CRGs) in lung tissues from Donate after Brain Death (DBD) and Donate after Cardiac-Dead (DCD) with the help of the R data package. Our study involved analyzing protein interactions, investigating gene correlations, and conducting functional enrichment. Key genes were identified using a range of machine learning techniques.

Results: Analysis of GEO datasets, which included 24 DBD and 11 DCD samples, identified two critical complement-related genes (CRGs): UNC119, FBLN2, and BCHE. The combination of these genes resulted in an Area Under the Curve (AUC) of 0.993, effectively distinguishing between lung tissues from DBD and DCD. Additionally, nomogram, decision curve, and calibration curve analyses confirmed their diagnostic efficacy.

Conclusion: We created a prognostic model for lung tissues from DBD and DCD donors using two pivotal Complement-Related Genes (CRGs), which showed impressive predictive power. These insights deepen our comprehension of the complement system's role in DBD and DCD lung tissues.

Cite this article: Liping L, Jiangfa L. Identification of complement-related genes in lung tissue after brain death and cardiac death based on machine learning. J Clin Med Images Case Rep. 2025; 5(1): 1772.

Introduction

The effects of brain death and cardiac death on lung tissue were different. Death is usually accompanied by neurogenic pulmonary edema and systemic inflammatory response, which is caused by the excessive release of norepinephrine from sympathetic nerve stimulation, resulting in increased permeability of pulmonary microvessels, resulting in acute pulmonary edema [1,2]. At the same time, people with brain death are often supported by mechanical ventilation, which can maintain oxygen supply while avoiding organ ischemia [3]. However, after cardiac death, rapid circulatory arrest leads to ischemia-reperfusion injury and rapid hypoxia of lung tissue, which leads to a

range of cellular damage and metabolic dysfunction, increasing the risk of pulmonary edema and inflammatory responses. In this case, donated lungs may not be as stable in texture as those obtained at brain death, and recovery after transplantation is difficult due to unmanageable ischemic time [4-6]. Therefore, understanding the different effects of the two on lung tissue is critical to optimizing strategies for the acquisition and preservation of transplanted organs.

The complement system is an important part of the innate immune response and plays multiple roles in inflammatory responses and tissue damage [7]. It consists of a series of plasma proteins that are activated in response to pathogen invasion or tissue damage, initiating expansion responses via the classical, bypass, and lectin pathways. Complement activation leads to increased cell lysis and phagocytosis, in which C3 and C5 complement components play a key role in regulating the inflammatory response by attracting neutrophils and other immune cells to the site of injury or infection. In addition, the complement system can directly cause pathogen destruction through the formation of Membrane Attack Complexes (MAC). However, overactivation or improper regulation of the complement system can lead to excessive inflammatory responses and tissue damage, exacerbating pathological damage and triggering diseases such as autoimmune disease and transplant rejection [8]. Therefore, the complement system plays a double-edged role in maintaining the balance between immune defense and damage repair.

In the pathological process of brain death and heart death, the activation and regulation of complement system show obvious differences. Activation of the complement system in brain-dead states is often strongly associated with Systemic Inflammatory Response Syndrome (SIRS), which may be due to significant changes in norepinephrine/epinephrine levels triggered by brain death. At this time, complement activation may exacerbate the inflammatory response, resulting in increased damage to lung and other tissues [9]. However, because the heart still keeps beating, complement-mediated damage is relatively better controlled. In contrast, after cardiac death, lung tissue is at risk of ischemia-reperfusion injury due to the rapid cessation of blood circulation, and the complement system is dramatically activated during reperfusion, resulting in a massive release of inflammatory cytokines. In this case, the complement system may further exacerbate reperfusion-related tissue damage and cell necrosis [10]. Therefore, the activation pathways of the complement system and its contribution to tissue damage in the two states of death show distinct patterns, and this difference needs to be deeply understood and managed during organ protection and transplantation.

Materials and methods

Patients and datasets

The transcriptomic analysis of lung specimens from DBD and DCD donors utilized the GSE18995 dataset, obtained from the GEO database, which included 24 DBD and 11 DCD samples. The flowchart illustrating this study is presented in (Figure 1).

Expression of DEGs and CRGs between DBD and DCD

We used the R package "limma" to perform a differential expression analysis on the processed data from the GEO database, identifying Differentially Expressed Genes (DEGs) between DBD and DCD. The results were visualized with volcano and heatmap plots using the "ggplot2" and "heatmap" R packages, with

the criterion for selection being an adjusted $P<0.05$. The intersection of DEGs ($|\log CF| >1$ and adjusted $P<0.05$) related to CRGs was identified using the "VennDiagram" R package and defined as DE-CRGs for further analysis. A boxplot for the DE-RGs was created using the "ggpubr" R package. A Venn diagram was used to pinpoint CRGs within DRGs between DBD and DCD samples. Subsequently, multivariate analysis was conducted to identify key DE-CRGs.

Correlation analysis and protein–protein interaction (PPI) network construction

The R package "heatmap" was used to generate heatmaps for 15 DE-CRGs. PPI networks for these DE-CRGs were constructed using the STRING database (https://string-db.org/). Furthermore, we employed the "ggplot2" package (version 3.3.6) in R to conduct pairwise correlation analysis of the dataset variables, visualizing the results in a heatmap.

Gene Ontology (GO) and KEGG pathway enrichment analysis

To analyze the biological function of genes, we employed the "clusterProfiler" package in R, which facilitated the enrichment analysis of Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways [11-13]. The GO annotation encompassed three domains: Biological Processes (BP), Cellular Components (CC), and Molecular Functions (MF).

Machine learning methods to identify key genes

The "glmnet" package in R was used to conduct Least Absolute Shrinkage and Selection Operator (LASSO) regression on the chosen linear model, effectively reducing data dimensionality while retaining essential variables [14]. Furthermore, we integrated Random Forest (RF) algorithms for optimal gene selection. RF, a regression tree technique, leverages bootstrap aggregation and randomization of predictors to achieve high prediction accuracy, implemented via the "randomForest" package in R [17]. The genes identified by these three machine learning methods were intersected to establish the final key genes.

The Establishment of the the DE-CRGs diagnostic Model

In developing the DE-CRGs diagnostic model, we utilized the GSE18995 dataset. Initially, we constructed a nomogram model using the "rms" package in R to predict the likelihood of DBD occurrence. Following this, we employed the "pROC" package [18] to assess the Area Under the Curve (AUC), along with the specificity and sensitivity of the diagnostic value for the marker genes through time-dependent ROC analysis. Each central gene was assigned a score, which was then aggregated to produce a total score.

Statistical analyses

Continuous variables are reported as mean ± standard deviation. The student's t-test was used for comparisons between two groups, while the Wilcoxon rank-sum test was applied to non-normally distributed variables. A p-value of less than 0.05 was considered statistically significant. The symbols *, **, and *** indicate p-values of less than 0.05, 0.01, and 0.001, respectively. All statistical analyses were conducted using R software (version 4.2.1).

Results

Identification of DE-CRGs associated with DBD

Using the "limma" package, we identified 3,868 DEGs (adj. p<0.05) from the GSE18995 dataset, which included 24 DBD and 11 DCD samples. Among these, 86 genes were up-regulated, and 109 were down-regulated. The volcano plot illustrating the differentially expressed genes is presented in (Figure 2A), while (Figure 2B) displays the heatmap of the top 50 differential genes between DBD and DCD. Additionally, 596 CRGs [19] overlapped with the 195 DEGs(($|\log CF| >1$ and adj. p<0.05), revealing 15 DE-CRGs with significant differences between the DBD and DCD (Figure 2C). Eight DE-CRGs (TAC, CXCR, SERPINE, CXCL8, PTX3, CCL2, IL6, and SELE) were high expression and seven DE-CRGs (C1QTNF2, UNC119, F13A1, CXCL12, FBLN2, BCHE, and IGHG1) were low expression in DBD (Figure 2D), and the heatmap of these 15 DE-CRGs is shown in (Figure 2E).

A PPI analysis was conducted using STRING to examine potential interactions among the 15 Differentially Expressed Complement-Related Genes (DE-CRGs), as shown in (Figure 3A). The correlations among these 15 DE-CRGs are presented in Figure 3B. DE-CRGs were found to be related to response to Lipopolysaccharide (LPS), inflammasome complex, cysteine-type endopeptidase activity, and cytokine receptor binding, among other pathways, as revealed by GO enrichment analysis (Figure 3C). KEGG pathway analysis showed involvement in lipid and atherosclerosis and non-alcoholic fatty liver disease (Figure 3C).

Identification of diagnostic marker genes for DBD

Considering the individual complexity and heterogeneity of DBD and DCD patients, candidate CRGs were identified from 15 DE-CRGs using RF and LASSO validated machine learning models to assist in predicting DBD diagnosis (Figures 4A & B). Out of the 15 DE-CRGs analyzed, three were successfully identified (refer to Figure 4C).

Evaluation of the diagnostic marker genes

A nomogram model was created to differentiate lung tissue from DBD and DCD, incorporating three central genes: UNC119, FBLN2, and BCHE (Figure 5A). The numerical values of each biomarker in the nomogram were used to distinguish between the two tissue types, and a calibration curve showed a strong correlation between predicted and actual probabilities (Figure 5B). Decision Curve Analysis (DCA) indicated that the net benefit from this model was significantly greater than zero, underscoring its considerable accuracy and utility in clinical decision-making (Figure 5C). ROC curve analysis demonstrated that the combined features of the three key genes performed well in distinguishing between DBD and DCD lung tissues (AUC=0.936, Figure 5D), with individual predictive ROC results for each gene closed to 0.90 (Figure 5E). These results suggest that the model based on these three marker genes has strong predictive efficacy for differentiating between lung tissues from DBD and DCD.
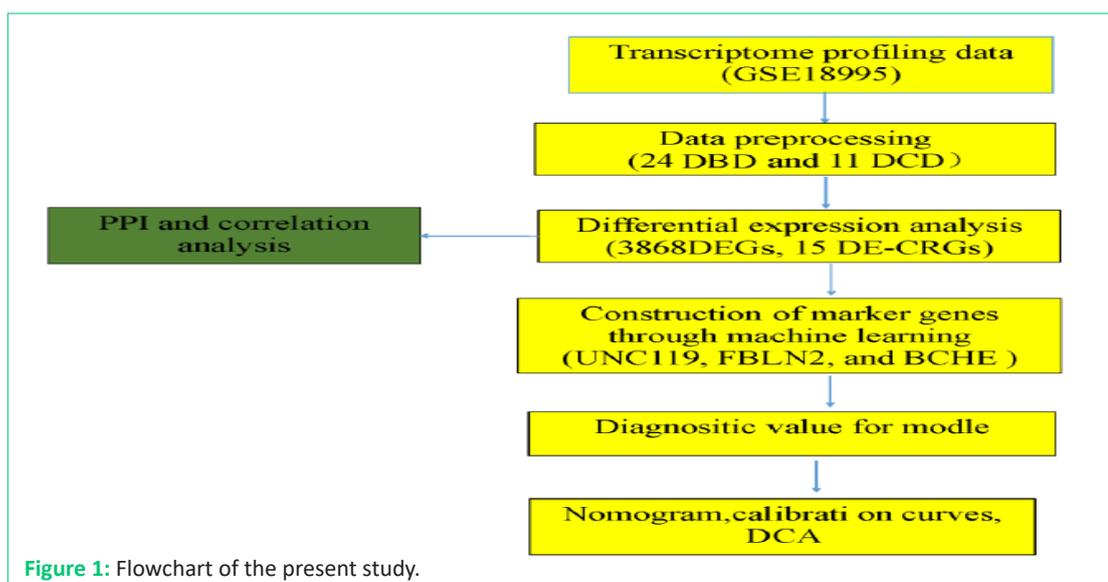
Figure 1: Flowchart of the present study.

Discussion

Brain Death (BD) donors are the main organ source for lung transplantation [20]. The death process can induce acute lung injury and aggravate lung ischemia-reperfusion injury [15]. While many immune mechanisms have been shown to stimulate donor organ damage and predispose grafts to poor outcomes, activation of the complement system has been shown to play a central role in BD-related graft damage. In both preclinical and clinical studies of heart and kidney transplants, grafts from BD donors were not only associated with more severe Ischemia-Reperfusion Injury (IRI), but posttransplant pathology was also significantly associated with levels of complement activation [21,22].

Complement is a major component of the immune system [23]. It is heavily involved in the body's defense function and its own immune regulation, and its overactivation plays a prim-

ing and amplifying effect in the initial stage of inflammatory response. The complement system is a particularly important upstream sensor and effector system, and the activation products of complement are important regulators in inflammatory and immune processes. As an important component of the immune system, complement is heavily involved in the body's defense response and its own immune regulation, and its overactivation plays an initiating and amplifying effect in the initial stage of inflammatory response [23]. Many causes of ALI have been experimentally demonstrated to be involved in the activation of the complement system [24].

In this study, we investigated the potential role of CRGs in lung tissue of DBD and DCD, identified potential key genes. We retrieved lung tissue samples from DBD and DCD patients from the GEO database for statistical analysis to identify DEGs, ultimately identifying 15 DEGs associated with complement pathways. These results imply that CRGs may play a role in the
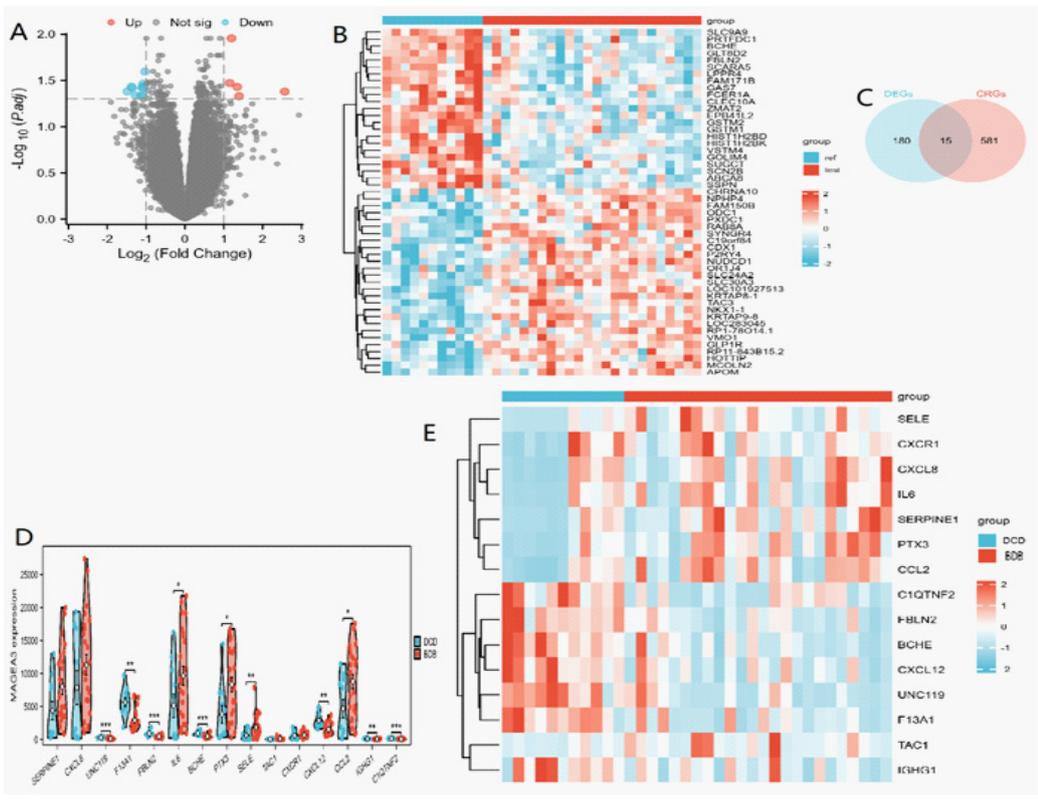
Figure 2: Identification of DE-CRGs. **(A)** A plot depicting the differential expression of genes between Brain-Dead (DBD) and Cardiac-Dead (DCD) donors. **(B)** Display of heatmaps for the most significant 50 genes. **(C)** Diagrams of Venn layout reveal commonalities between Differentially Expressed Genes (DEGs) and Complement-Related Genes (CRGs). **(D)** Presentation of fifteen Differentially Expressed pyroptosis-Related Genes (DE-CRGs) with accompanying boxplots showing variations in expression levels between DBD and DCD scenarios. **(E)** A heatmap is shown to illustrate the expression patterns for these fifteen DE-CRGs. Statistical significance is indicated as follows: *p<0.05; **p<0.01; ***p<0.001., CRGs: Complement-Related Genes; DBD: Donate after Brain-Dead; DCD: Donate after Cardiac-Dead.
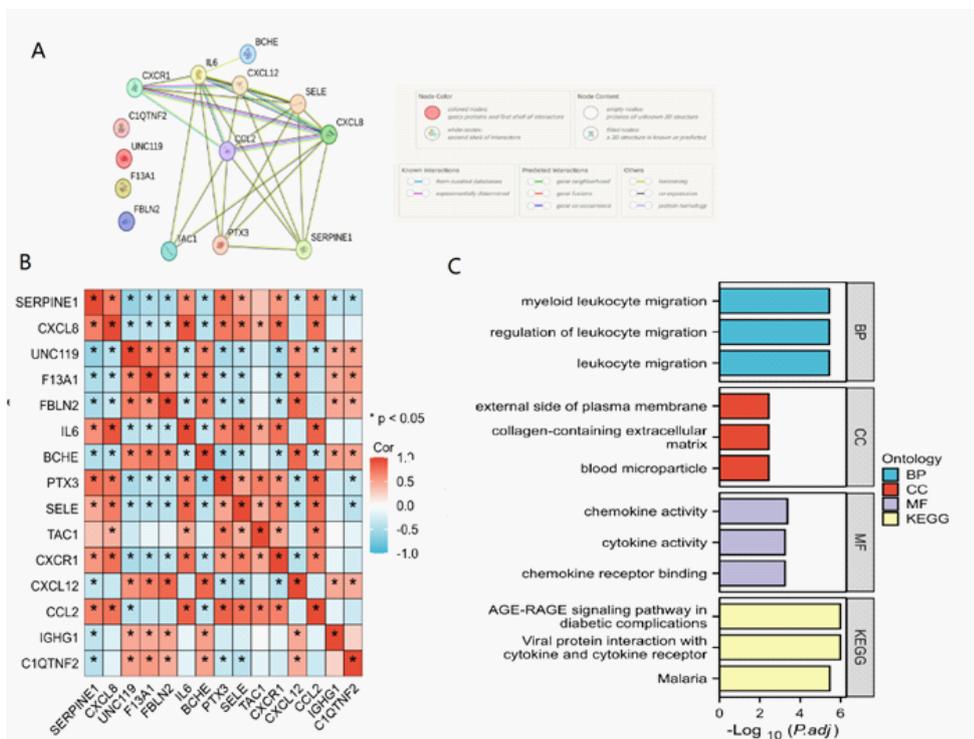
Figure 3: PPI, GO, and KEGG Analysis of 15 DE-CRGs. **(A)** This panel displays a network diagram highlighting the relationships among the fifteen DE-CRGs. **(B)** A correlation analysis for these fifteen DE-CRGs is visualized, with orange indicating positive correlations and blue denoting negative correlations. **(C)** Functional analysis was performed on ten of these DE-CRGs using Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) methodologies. Statistical significance for the analyses is noted with *p<0.05. DE-CRGs: Differentially expressed pyroptosis-related genes.
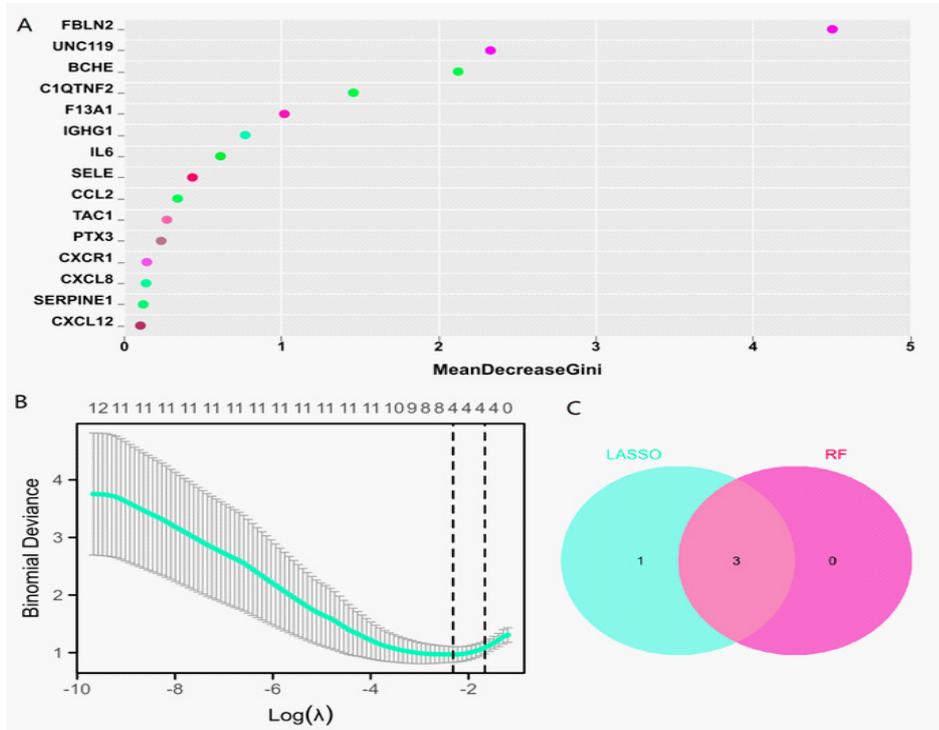
Figure 4: Machine Learning Identification of Diagnostic Marker Genes. **(A)** In the Random Forest (RF) analysis, three genes were identified having Gini values exceeding 2. **(B)** Analysis of coefficients was performed using the Least Absolute Shrinkage and Selection Operator (LASSO) method. **(C)** A Venn diagram illustrates the shared genes identified by both machine learning algorithms, namely RF and LASSO.
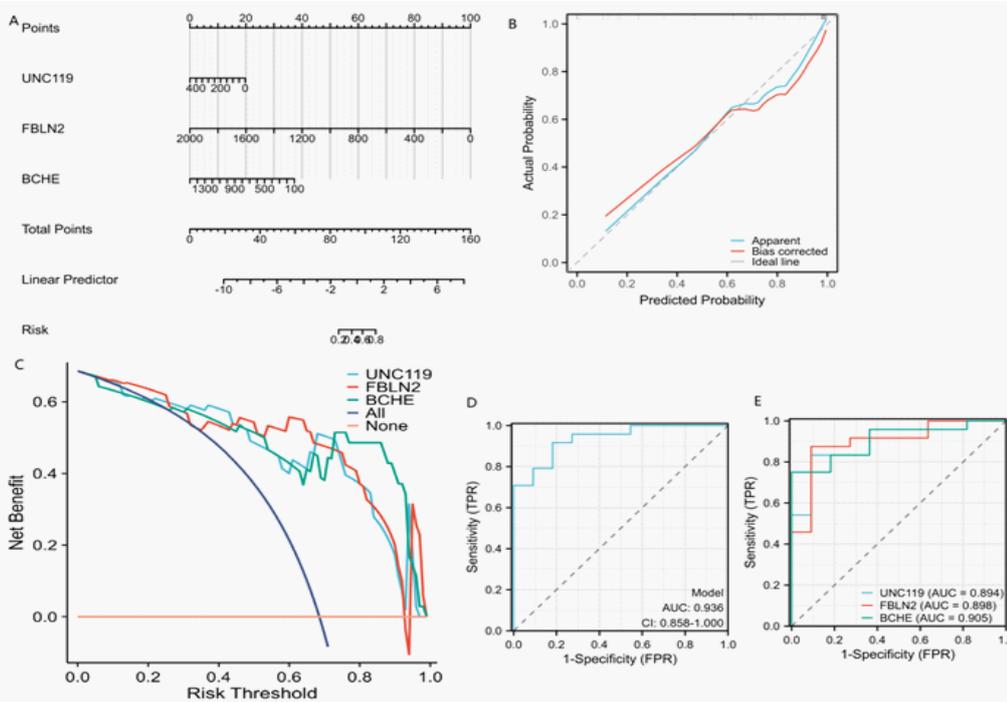
Figure 5: Establishment of Marker Gene Diagnostic Model. **(A)** A nomogram has been constructed for the marker genes to facilitate the prediction of clinical outcomes. **(B)** A calibration curve was developed to evaluate the accuracy of the nomogram in comparison to actual outcomes. **(C)** The effectiveness of the nomogram model's predictions was demonstrated using Decision Curve Analysis (DCA), highlighting the clinical utility by comparing the net benefits across a range of threshold probabilities against two reference strategies: "treat all" and "treat none." **(D)** The combined predictive power of the three key genes was quantified by a Receiver Operating Characteristic (ROC) curve, achieving an Area Under the Curve (AUC) of 0.936, with a confidence interval (95% CI) ranging from 0.976 to 1.0, indicating a high level of diagnostic accuracy. **(E)** Additionally, the ROC results for the individual key genes—UNC119, FBLN2, and BCHE—were reported with AUC values of 0.894, 0.898, and 0.905, respectively, underlining each gene's significant predictive capability. This detailed analysis and comparison with "all" or "none" treatment strategies elucidate the substantial benefits and clinical applicability of the nomogram, supporting its potential use in personalized medicine to make informed treatment decisions based on genetic profiling.

progression of lung injury. Our correlation analysis showed that the identified DE-CRGs were closely interconnected; however, some exhibited no clear correlation at the protein level, emphasizing the heterogeneity in CRG interactions at both gene and protein levels.

The important role of DE-CRGs in response to myeloid leukocyte migration, regulation of leukocyte migration, external side of plasma membrane, collagen-containing extracellular matrix, chemokine activity chemokine receptor binding, viral protein interaction with cytokine and cytokine receptor, AGE-RAGE signaling pathway in diabetic complications, and malaria was revealed by GO and KEGG enrichment analyses.

Analyses using LASSO, and RF of the 15 DE-CRGs identified three key genes—UNC119, FBLN2, and BCHE—that effectively differentiate lung tissue from DBD and DCD, achieving an AUC of 0.936. The nomogram model, supported by calibration curves and Decision Curve Analysis (DCA), demonstrated robust predictive capability and significant clinical relevance.

UNC119 is a conserved intracellular protein that mainly plays a role in the nervous system, especially in neurons [25], and is involved in a variety of biological processes such as cell signal transduction, cytoskeleton regulation and membrane transport. UNC119 binds to some intracellular small molecules (such as small GTPase), regulates their activity, and plays an important role in the development and function of neurons [26]. UNC119 may regulate the function of immune cells by affecting intracellular signal transduction pathway, thereby indirectly affecting the activation and regulation of complement [27]. In immune response, UNC119 may be involved in the interaction between immune cells and complement components, affecting their ability to recognize and clear pathogens. UNC119 may play a regulatory role in the development and function of immune cells, thereby affecting complement-related immune response. FBLN2 (Fibulin-2) is an extracellular matrix protein, belonging to the fibronectin family, which plays an important role in tissue development, repair and maintenance. FBLN2 plays a key role in a variety of biological processes, including cell adhesion, signaling, and the formation of tissue structure [28-30]. FBLN2 may affect the function of complement system by influencing the structure of extracellular matrix and altering the deposition and activation process of complement components. FBLN2 is expressed in certain immune cells, such as macrophages and lymphocytes, and may influence their response to complement activation by regulating the adhesion and migration of these cells. Butyrylcholinesterase (BCHE) is an enzyme in blood plasma that is primarily responsible for the hydrolysis of cholinesterase compounds [31] in the body, BCHE plays an important role in neurotransmission, drug metabolism, detoxification and other processes. Similar to Acetylcholinesterase (AChE), BCHE acts mainly in some tissues close to neurons, including blood and liver, but it has a broader substrate specificity. The main physiological function of BCHE is to hydrolyze Butyrylcholine and other related ester compounds, which play an important role in the regulation of neurotransmission, although its role is relatively limited compared with AChE. The activation of the complement system usually triggers an inflammatory response, and BCHE may play a supporting role in this process, by hydrolyzing and clearing some inflammatory mediators, alleviating the intensity of the inflammatory response. This also hints at its potential role in regulating complement-mediated inflammation [32].

Our study has some limitations. Firstly, we conducted genetic analysis on data sourced from the GEO database, which may introduce certain biases. Secondly, the sample size was relatively small, which could impact the generalizability of the findings. Finally, we did not conduct additional tests to verify the expression of these genes.

Conclusion

We initially identified two significant genes that, in combination, accurately differentiate between lung tissue from DBD and DCD.

Declarations

References

1. Migdady I, Rae-Grant A, Greer DM. Brain death evaluation during the pandemic. Neurology. 2020; 95: 693-694.

2. Snell GI, Levvey BJ, Levin K, Paraskeva M, Westall G. Donation after Brain Death versus Donation after Circulatory Death: Lung Donor Management Issues. Semin Respir Crit Care Med. 2018; 39: 138-147.

3. Belhaj A, Dewachter L, Hupkens E, Remmelink M, Galanti L, Rorive S et al. Tacrolimus Prevents Mechanical and Humoral Alterations in Brain Death-induced Lung Injury in Pigs. Am J Respir Crit Care Med. 2022; 206: 584-595.

4. Ceulemans LJ, Inci I, Van Raemdonck D. Lung donation after circulatory death[J]. Current opinion in organ transplantation.

2019; 24: 288-296.

5. Van Raemdonck D, Ceulemans LJ, Neyrinck A, Levvey B, Snell GI. Donation After Circulatory Death in lung transplantation. Thoracic surgery clinics. 2022; 32: 153-165.

6. Santos P, Teixeira PJZ, Moraes Neto DM, Cypel M. Donation after circulatory death and lung transplantation. Jornal brasileiro de pneumologia : publicacao oficial da Sociedade Brasileira de Pneumologia e Tisilogia. 2022; 48: e20210369.

7. Conigliaro P, Triggianese P, Ballanti E, Perricone C, Perricone R, Chimenti MS. Complement, infection, and autoimmunity. Current opinion in rheumatology. 2019; 31: 532-541.

8. Santarsiero D, Aiello S. The Complement System in Kidney Transplantation. Cells. 2023; 12.

9. Thorgersen EB, Barratt-Due A, Haugaa H, Harboe M, Pischke SE, Nilsson PH et al. The Role of Complement in Liver Injury, Regeneration, and Transplantation. Hepatology, 2019; 70: 725-736.

10. Ali HA, Pavlisko EN, Snyder LD, Frank M, Palmer SM. Complement system in lung transplantation. Clin Transplant. 2018; 32: e13208.

11. Kanehisa M, Furumichi M, Sato Y, Kawashima M, Ishiguro-Watanabe M. KEGG for taxonomy-based analysis of pathways and genomes. Nucleic acids research. 2023; 51: D587-D592.

12. Kanehisa M, Goto S. KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 2000; 28.

13. Kanehisa M. Toward understanding the origin and evolution of cellular organisms. Protein Sci. 2019: 28.

14. van Egmond MB, Spini G, van der Galien O, A IJ, Veugen T, Kraaij W et al. Privacy-preserving dataset combination and Lasso regression for healthcare predictions. BMC Med Inform Decis Mak. 2021; 21: 266.

15. Joshi P, Vedhanayagam M, Ramesh R. An Ensembled SVM Based Approach for Predicting Adverse Drug Reactions. Current Bioinformatics. 2021; 16: 422-432.

16. Abinash MJ, Vasudevan V. Boundaries tuned support vector machine (BT-SVM) classifier for cancer prediction from gene selection. Comput Methods Biomech Biomed Engin. 2022; 25: 794-807.

17. Rigatti SJ. Random Forest. J Insur Med. 2017; 47: 31-39.

18. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics. 2011; 12: 77.

19. Wu J, Zhu Y, Luo M, Li L. Comprehensive Analysis of Pyroptosis-Related Genes and Tumor Microenvironment Infiltration Characterization in Breast Cancer. Front Immunol. 2021; 12: 748221.

20. Hoang V, Medrek SK, Pendurthi M, Kao C, Parulekar AD. Improving Donor Lung Management and Recipient Selection in Lung Transplantation. Am J Respir Crit Care Med. 2017; 196: 782-784.

21. Atkinson C, Floerchinger B, Qiao F, Casey S, Williamson T, Moseley E et al. Donor brain death exacerbates complement-dependent ischemia/reperfusion injury in transplanted hearts. Circulation. 2013; 127: 1290-1299.

22. Damman J, Nijboer WN, Schuurs TA, Leuvenink HG, Morariu AM, Tullius SG, et al. Local renal complement C3 induction by donor brain death is associated with reduced renal allograft function after transplantation. Nephrology, dialysis, transplantation: official publication of the European Dialysis and Transplant Association - European Renal Association. 2011; 26: 2345-2354.

23. Holers VM. The spectrum of complement alternative pathway-mediated diseases. Immunological reviews. 2008; 223: 300-316.

24. Petersen B, Busch T, Gaertner J, Haitsma JJ, Krabbendam S, Ebsen M, et al. Complement activation contributes to ventilator-induced lung injury in rats. Journal of physiology and pharmacology: an official journal of the Polish Physiological Society. 2016; 67: 911-918.

25. Zhang H, Constantine R, Vorobiev S, Chen Y, Seetharaman J, Huang YJ, et al. UNC119 is required for G protein trafficking in sensory neurons. Nature neuroscience. 2011; 14: 874-880.

26. Frederick JM, Hanke-Gogokhia C, Ying G, Baehr W. Diffuse or hitch a ride: how photoreceptor lipidated proteins get from here to there. Biological chemistry. 2020; 401: 573-584.

27. Gorska MM, Alam R. Consequences of a mutation in the UNC119 gene for T cell function in idiopathic CD4 lymphopenia. Current allergy and asthma reports. 2012; 12: 396-401.

28. de Vega S, Iwamoto T, Yamada Y. Fibulins: multiple roles in matrix structures and tissue functions. Cellular and molecular life sciences: CMLS. 2009; 66: 1890-1902.

29. Argraves WS, Greene LM, Cooley MA, Gallagher WM. Fibulins: physiological and disease perspectives. EMBO reports. 2003; 4: 1127-1131.

30. Timpl R, Sasaki T, Kostka G, Chu ML. Fibulins: a versatile family of extracellular matrix proteins[J]. Nature reviews Molecular cell biology. 2003; 4: 479-489.

31. Brimijoin S, Chen VP, Pang YP, Geng L, Gao Y. Physiological roles for butyrylcholinesterase: A BChE-ghrelin axis. Chemico-biological interactions. 2016; 259: 271-275.

32. Liu J, Tian T, Liu X, Cui Z. BCHE as a Prognostic Biomarker in Endometrial Cancer and Its Correlation with Immunity. Journal of immunology research. 2022; 2022: 6051092.