



## OPEN ACCESS

## EDITED BY

Henggui Zhang,  
The University of Manchester,  
United Kingdom

## REVIEWED BY

Gary C. Mouradian,  
Medical College of Wisconsin,  
United States  
Christopher G. Wilson,  
Loma Linda University, United States

## \*CORRESPONDENCE

Feng Jiang,  
✉ dxyjiang@163.com  
Chuyan Wu,  
✉ chuyan\_w@hotmail.com  
Fei Luo,  
✉ fduluofei@163.com

## SPECIALTY SECTION

This article was submitted to  
Computational Physiology and Medicine,  
a section of the journal  
Frontiers in Physiology

RECEIVED 30 October 2022

ACCEPTED 23 December 2022

PUBLISHED 09 January 2023

## CITATION

Tao Z, Mao Y, Hu Y, Tang X, Wang J,  
Zeng N, Bao Y, Luo F, Wu C and Jiang F  
(2023), Identification and immunological  
characterization of endoplasmic reticulum  
stress-related molecular subtypes in  
bronchopulmonary dysplasia based on  
machine learning.  
*Front. Physiol.* 13:1084650.  
doi: 10.3389/fphys.2022.1084650

## COPYRIGHT

© 2023 Tao, Mao, Hu, Tang, Wang, Zeng,  
Bao, Luo, Wu and Jiang. This is an open-  
access article distributed under the terms  
of the [Creative Commons Attribution  
License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or  
reproduction in other forums is permitted,  
provided the original author(s) and the  
copyright owner(s) are credited and that  
the original publication in this journal is  
cited, in accordance with accepted  
academic practice. No use, distribution or  
reproduction is permitted which does not  
comply with these terms.

# Identification and immunological characterization of endoplasmic reticulum stress-related molecular subtypes in bronchopulmonary dysplasia based on machine learning

Ziyu Tao<sup>1</sup>, Yan Mao<sup>2</sup>, Yifang Hu<sup>3</sup>, Xinfang Tang<sup>4</sup>, Jimei Wang<sup>5</sup>,  
Ni Zeng<sup>6</sup>, Yunlei Bao<sup>5</sup>, Fei Luo<sup>5\*</sup>, Chuyan Wu<sup>7\*</sup> and Feng Jiang<sup>5\*</sup>

<sup>1</sup>Department of Ultrasound, Obstetrics and Gynecology Hospital of Fudan University, Shanghai, China,

<sup>2</sup>Department of Pediatrics, The First Affiliated Hospital of Nanjing Medical University, Nanjing, China,

<sup>3</sup>Department of Geriatric Endocrinology, The First Affiliated Hospital of Nanjing Medical University, Nanjing, China,

<sup>4</sup>Department of Nephrology, The Affiliated Lianyungang Oriental Hospital of Xuzhou Medical University, The Affiliated Lianyungang Oriental Hospital of Kangda College of Nanjing Medical University, The

Affiliated Lianyungang Oriental Hospital of Bengbu Medical College, Lianyungang, China, <sup>5</sup>Department of

Neonatology, Obstetrics and Gynecology Hospital of Fudan University, Shanghai, China, <sup>6</sup>Department of

Dermatology, Affiliated Hospital of Zunyi Medical University, Zunyi, China, <sup>7</sup>Department of Rehabilitation

Medicine, The First Affiliated Hospital of Nanjing Medical University, Nanjing, China

**Introduction:** Bronchopulmonary dysplasia (BPD) is a life-threatening lung illness that affects premature infants and has a high incidence and mortality. Using interpretable machine learning, we aimed to investigate the involvement of endoplasmic reticulum (ER) stress-related genes (ERSGs) in BPD patients.

**Methods:** We evaluated the expression profiles of endoplasmic reticulum stress-related genes and immune features in bronchopulmonary dysplasia using the GSE32472 dataset. The endoplasmic reticulum stress-related gene-based molecular clusters and associated immune cell infiltration were studied using 62 bronchopulmonary dysplasia samples. Cluster-specific differentially expressed genes (DEGs) were identified utilizing the WGCNA technique. The optimum machine model was applied after comparing its performance with that of the generalized linear model, the extreme Gradient Boosting, the support vector machine (SVM) model, and the random forest model. Validation of the prediction efficiency was done by the use of a calibration curve, nomogram, decision curve analysis, and an external data set.

**Results:** The bronchopulmonary dysplasia samples were compared to the control samples, and the dysregulated endoplasmic reticulum stress-related genes and activated immunological responses were analyzed. In bronchopulmonary dysplasia, two distinct molecular clusters associated with endoplasmic reticulum stress were identified. The analysis of immune cell infiltration indicated a considerable difference in levels of immunity between the various clusters. As measured by residual and root mean square error, as well as the area under the curve, the support vector machine machine model showed the greatest discriminative capacity. In the end, an support vector machine model integrating five genes was developed, and its performance was shown to be excellent on an external validation dataset. The effectiveness in predicting bronchopulmonary

dysplasia subtypes was further established by decision curves, calibration curves, and nomogram analyses.

**Conclusion:** We developed a potential prediction model to assess the risk of endoplasmic reticulum stress subtypes and the clinical outcomes of bronchopulmonary dysplasia patients, and our work comprehensively revealed the complex association between endoplasmic reticulum stress and bronchopulmonary dysplasia.

#### KEYWORDS

bronchopulmonary dysplasia, machine learning, endoplasmic reticulum stress, immune infiltration, prediction model

## Introduction

Bronchopulmonary dysplasia (BPD) is a pulmonary condition that affects preterm newborns and is hallmarked by suppressed lung development and lung damage, requiring life-saving medical interventions like mechanical ventilation and oxygen treatment (Thebaud et al., 2019). Consequently, it is one of the conditions that is considered to be the most life-threatening in neonatal intensive care units (Bancalari and Jain, 2019). Even though there have been tremendous breakthroughs in maternal and neonatal care, the prevalence of BPD in preterm babies remains between 20% and 40% (Sillers et al., 2020). BPD may result in impaired lung development and permanent impairment (Gilfillan et al., 2021). BPD may also cause harm to the neurological system and other systemic organs (Shukla and Ambalavanan, 2021). However, the pathogenesis of BPD is poorly known, and there are few early diagnostic tests and precise prevention or therapy approaches for this disease.

The endoplasmic reticulum (ER) is essential for maintaining the intracellular environment's equilibrium (Zeeshan et al., 2016). Under persistent stress, the disturbance of ER homeostasis may result in ER stress, as demonstrated by alterations in the Ca<sup>2+</sup> levels in the cells and the overaccumulation of misfolded or unfolded proteins, which ultimately leads to numerous protein-folding diseases, including BPD (Adamopoulos et al., 2014; Leprivier et al., 2015). It is generally believed that prematurely born infants display oxidative stress shortly after delivery owing to their immature antioxidant defenses and the fast increase in oxygen tension caused by medical intervention (Bonadies et al., 2020). Numerous adaptive survival mechanisms are triggered in response to a sudden rise in oxidative stress, such as the UPR (a reaction to variations in chaperone function), resulting in increased ER stress and reduced protein production to facilitate the restoration of proteostasis (Hu et al., 2020; Kepp and Galluzzi, 2020). Both ER stress and oxidative stress are two connected conditions associated with a variety of lung diseases (Marciniak, 2017; Chen et al., 2018). Recent research indicates that ER stress is elevated in the lungs of numerous recognized BPD animal models (Pritchard et al., 2022). These investigations showed that ER stress may be an effective therapeutic target for BPD. However, the particular molecular processes of ER in controlling the course of BPD have yet to be completely explained and need more research.

We investigated the expression patterns of ER stress-related genes (ERSGs) between BPD and control samples, as well as the association between differentially expressed genes (DEGs), linked to ER stress immune features. Then, we grouped BPD patients into two ER stress-related clusters and analyzed the differences (variations) in immune cell infiltration between the two groups. The WGCNA approach was then used to detect cluster-specific DEGs, and enriched bioactivities

and pathways were determined based on these cluster-specific DEGs. Moreover, the analysis of multiple machine-learning approaches resulted in the development of a prediction model for identifying patients who have distinct molecular clusters. After that, we verified the nomogram, calibration curve, decision curve analysis (DCA), and the predictive model's performance, which offered some additional unique knowledge for predicting BPD clusters and relevant risk.

## Materials and methods

### Acquisition and preprocessing of data

Gene Expression Omnibus (GEO) was searched to acquire two microarray datasets (GSE32472 and GSE108756) associated with BPD. Microarray profiles of gene expression in neonates with BPD were reported by GSE32472 using blood samples taken around the 5th, 14th, and 28th days of life. For reliability, we chose 100 blood samples on around the 28th day, when a more precise diagnosis of BPD could be established, which mainly concluded 38 controls and 62 patients with BPD. GSE108756 included six control samples and five BPD samples and was used as an external validation set. To standardize the expression data, the quantile normalization function in the limma package of R was used. The ERSGs were obtained from two gene sets (GOBP response to endoplasmic reticulum stress and GOBP regulation of response to endoplasmic reticulum stress) of MSigDB.v2022 and 260 ERSGs were collected (Supplementary Table S1).

### Analysis of immune cell infiltration

To calculate the relative abundance of 22 different types of immune cells in each sample using the processed gene expression profiles, the CIBERSORT method and LM22 signature matrix were used (Newman et al., 2015). When calculating the inverse fold product probability value, CIBERSORT employs Monte Carlo sampling. We only counted immune cell fractions with *p*-values below .05 as accurate. The sum of the 22 immune cell proportions in each sample was 1.

### Analysis of the relationship between ERSGs and infiltrating immune cells

We examined correlation coefficients between ERSGs expression and the relative proportion of immune cells as a further demonstration of the connection between ERSGs and BPD-associated immune

features.  $p$ -values  $< .05$  for the spearman correlation coefficient indicated a statistically significant relationship. Lastly, the “correplot” R package was used to visually display the findings.

## Unsupervised clustering of BPD patients

We analyzed the differences in ERSGs between BPD and control groups, and the screening criteria were  $|\log_{2}FC| > 0$  and  $p < .001$  for genes with significant differences. We got 31 differentially expressed ERSGs. Using the expression patterns of 31 ERSGs as input, we executed an unsupervised clustering analysis utilizing the “ConsensusClusterPlus” R package (Wilkerson and Hayes, 2010), which used the k-means method and 1,000 interactions to classify the 62 BPD samples into distinct clusters. On the basis of the consensus matrix, the cumulative distribution function (CDF) curve (Boso et al., 2014), and the consistent cluster score ( $> .9$ ), we determined that  $k = 2$  was the maximum number of subtypes in our analysis and thoroughly evaluated the optimal cluster number.

## Gene set variation analysis (GSVA)

The “GSVA” package (Hanzelmann et al., 2013) in R was employed to perform enrichment analysis, which helped to reveal the variations in enriched gene sets across various ER stress clusters. From the MSigDB website database, we obtained the “c2.cp.kegg.symbols.gmt” and “c5.go.symbols.gmt” files to conduct additional GSVA analysis. By comparing GSVA scores across several ER stress clusters, the “limma” R program was used to determine the pathways and biological functions that were expressed differently. Outcomes in the GSVA score with a  $t$ -value greater than two were judged significant.

## Weighted gene co-expression network analysis (WGCNA)

The “WGCNA” package in R was employed to execute WGCNA and find co-expression modules (Langfelder and Horvath, 2008). To ensure the reliability of high-quality findings, the top 25% of genes with the greatest variation were subjected to WGCNA analyses. The optimal soft thresholding power was used to generate the weighted adjacency matrix, which was then used as the basis for a topological overlap matrix (TOM). Additionally, the TOM dissimilarity measure (1-TOM) predicated on the hierarchical clustering tree technique was used to create modules with a minimum module size of 100. A different color was chosen at random for each module. The global gene expression patterns in each module were denoted by the eigengene of the corresponding module. Modular significance (MS) demonstrated the connection between modules and illness conditions. The link between a gene and its clinical phenotypes was termed gene significance (GS).

## The development of a prediction model using various machine-learning techniques

We used the “caret” R package to develop machine learning models, which include an eXtreme Gradient Boosting (XGB), generalized linear model (GLM), support vector machine model

(SVM), and random forest model (RF), premised on two distinct ERSG clusters. RF is a machine learning ensemble method that predicts categorization or regression using several, unrelated decision trees (Rigatti, 2017). The SVM technique allows for the generation of a hyperplane in the characteristic space that has a hyperplane that maximizes the margin to discriminate between negative and positive examples (Giraldo et al., 2006). GLM is a variant of the linear regression technique used to examine the link between normal distributions of dependent data and continuous or categorical independent characteristics. XGB is a gradient-boosting-based ensemble of boosted trees that can quantitatively compare model complexity and classification error (Lee and Nelder, 2002). Cluster-specific DEGs were chosen as explanatory variables, whereas the response variables were distinct clusters. We classified 62 BPD samples at random into a training set (consisting of 70%) and a validation set (consisting of 30%). Machine learning models were run with default settings and evaluated by means of 5-fold cross-validation, and parameter optimization was done automatically by the caret package using grid search. To clearly explain the link between these four machine learning models and to compare their residual distributions and feature significance, the “DALEX” package was implemented. To display the area under ROC curves, the “pROC” (Robin et al., 2011) R program was used. As a result, the best machine learning model was selected, and the top five factors served as the key predictor genes for BPD.

## Development and verification of a nomogram model

With the aid of the “rms” R package, we developed a nomogram to analyze BPD cluster-related occurrence. There is a corresponding score assigned to each predictor, and this value is added together to get the “total score.” To evaluate the nomogram model’s prediction ability, we used the DCA (Vickers and Elkin, 2006) and calibration curve.

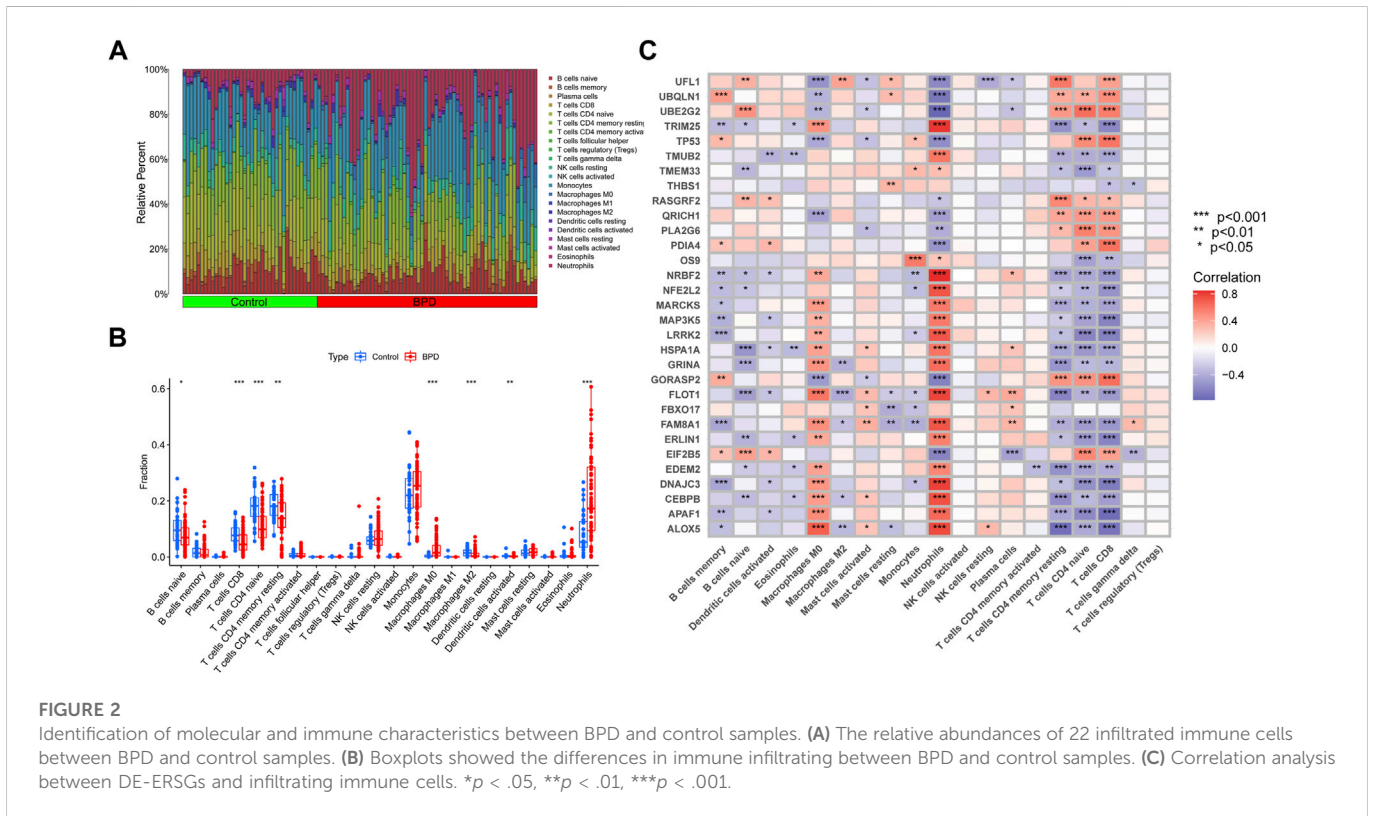
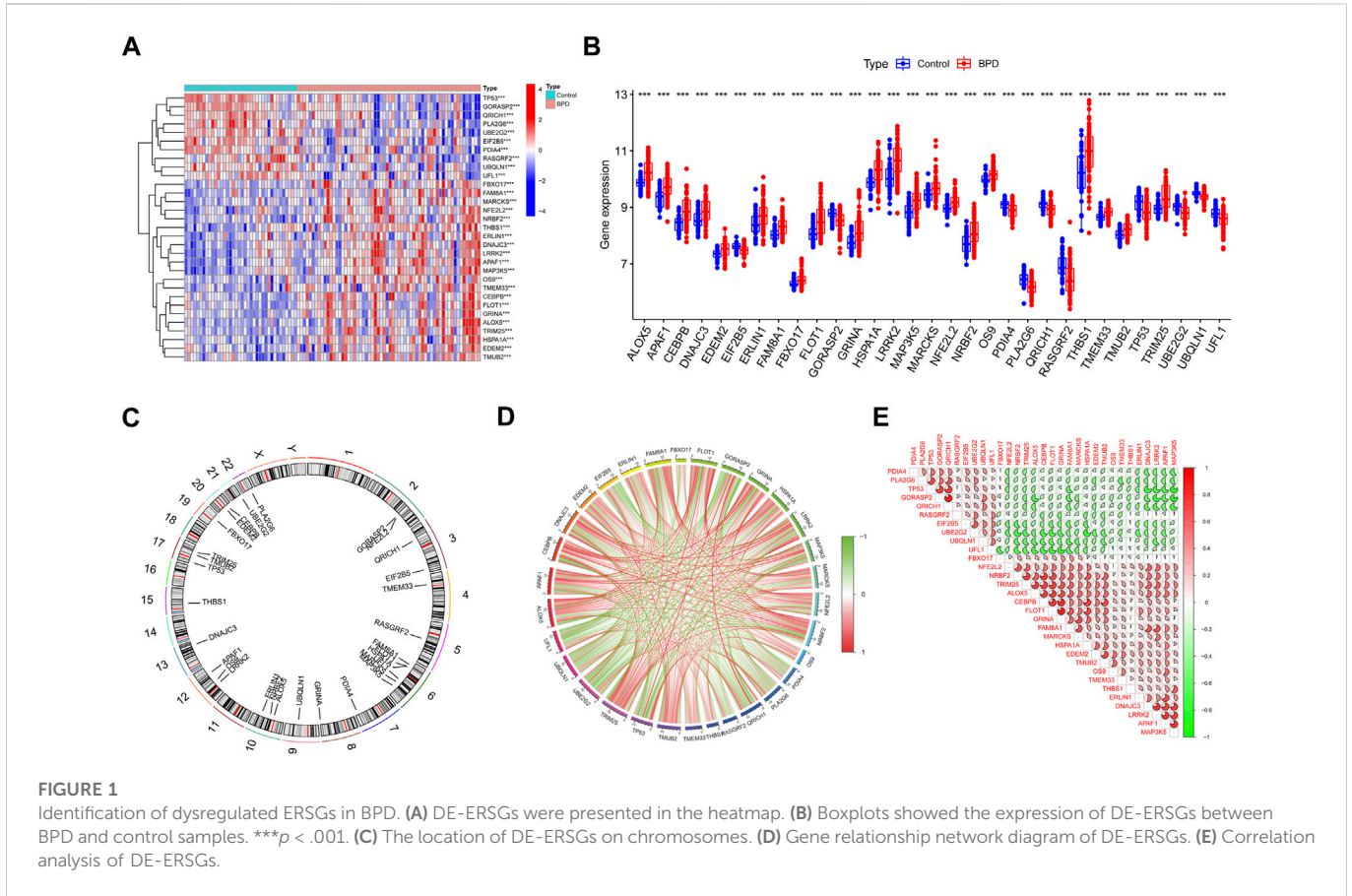
## Independent validation analysis

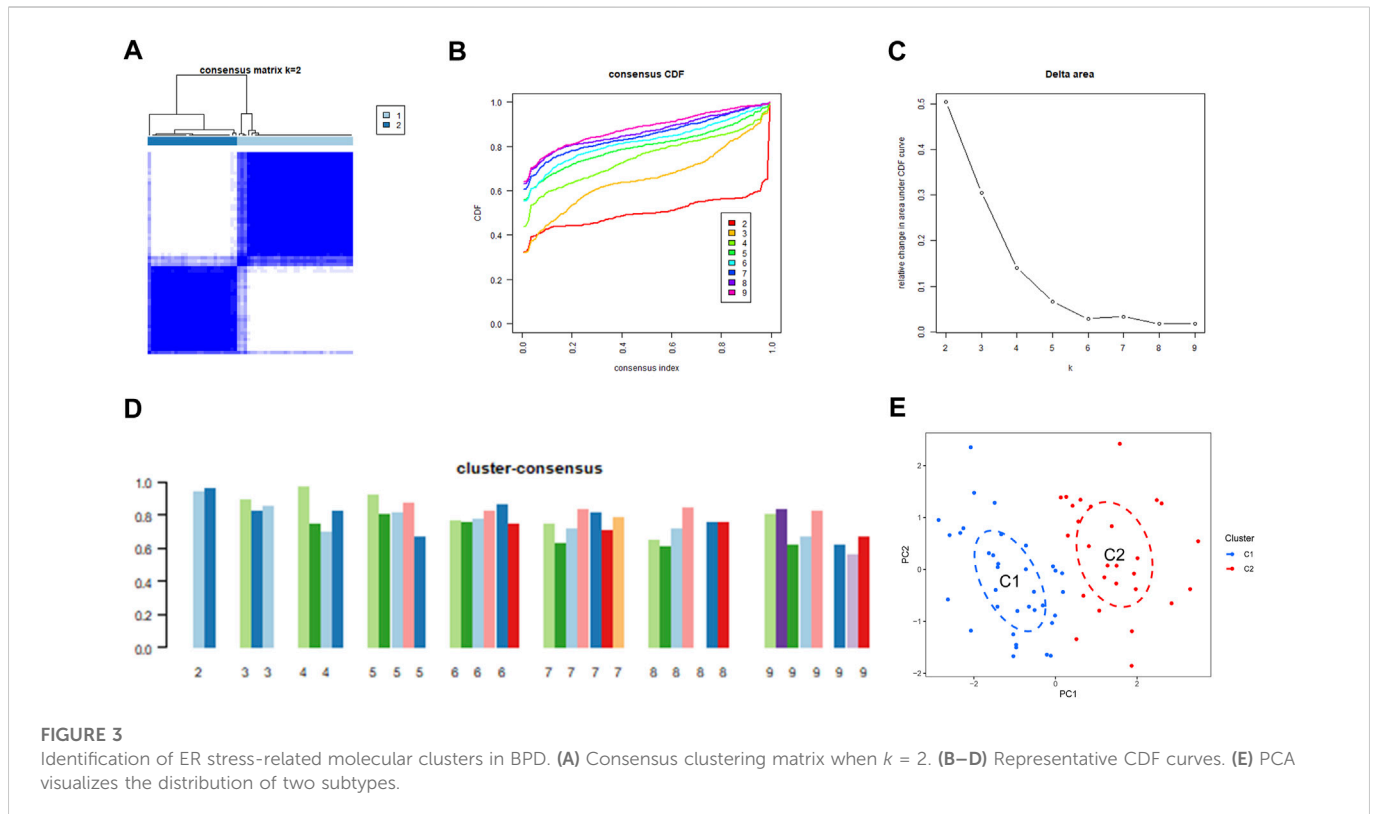
The ROC analyses were used to verify the predictive model’s capacity to differentiate between BPD and non-BPD controls using the data set GSE108756. The “pROC” R package was used to visualize ROC curves.

## Results

### Dysregulation of ERSGs

We initially extensively made a comparison of the expression patterns of 260 ERSGs across BPD and control samples based on the GSE32472 dataset, to elucidate the biological involvement of ERSGs in the onset and advancement of BPD. In total, 31 ERSGs were determined as the differentially expressed ERSGs (DE-ERSGs) according to the criteria of  $|\log_{2}FC| > 0$  and  $p < .001$  (Figures 1A–C). To further probe whether ERSGs played a crucial function in the onset and advancement of BPD, we conducted a correlation study across these DE-ERSGs. Many of the ERSGs presented strong





synergistic effects (Figure 1D). Additional evidence of the remarkable link between these DE-ERSGs was provided by the gene interaction network diagram (Figure 1E).

## Activation of the immune responses in BPD patients

Immune cell infiltration analysis using the CIBERSORT method demonstrated a difference in the percentages of 22 types of infiltrated immune cells between BPD and control samples, providing more evidence for the existence of immune system differences between the two groups (Figure 2A). The findings illustrated that BPD patients exhibited a greater infiltration level of monocytes, macrophages M0, neutrophils, naive CD4<sup>+</sup> T cell, and lower levels of dendritic cells activated, macrophages M2, resting memory CD4 T cells, and CD8 T cells (Figure 2B), pointing to immune system changes as a potential root cause of BPD. Nevertheless, a study of correlations found strong associations between ERSGs and neutrophils, macrophages, resting CD4 memory T cells, naive CD4 T cells, and CD8 T cells (Figure 2C). From these findings, ERSGs could perform a crucial function in controlling the molecular and immunological infiltration status of BPD patients.

## Determination of ERSGs clusters in BPD

We used a consensus clustering technique to classify the 62 samples according to the expression patterns of 31 DE-ERSGs, thus elucidating the ER stress-associated expression profiles in BPD. Cluster sizes were most consistent with a  $k$  value of 2 ( $k = 2$ ), whereas CDF curves showed little variation between consensus indices of .2 and .8 (Figures 3A–C).

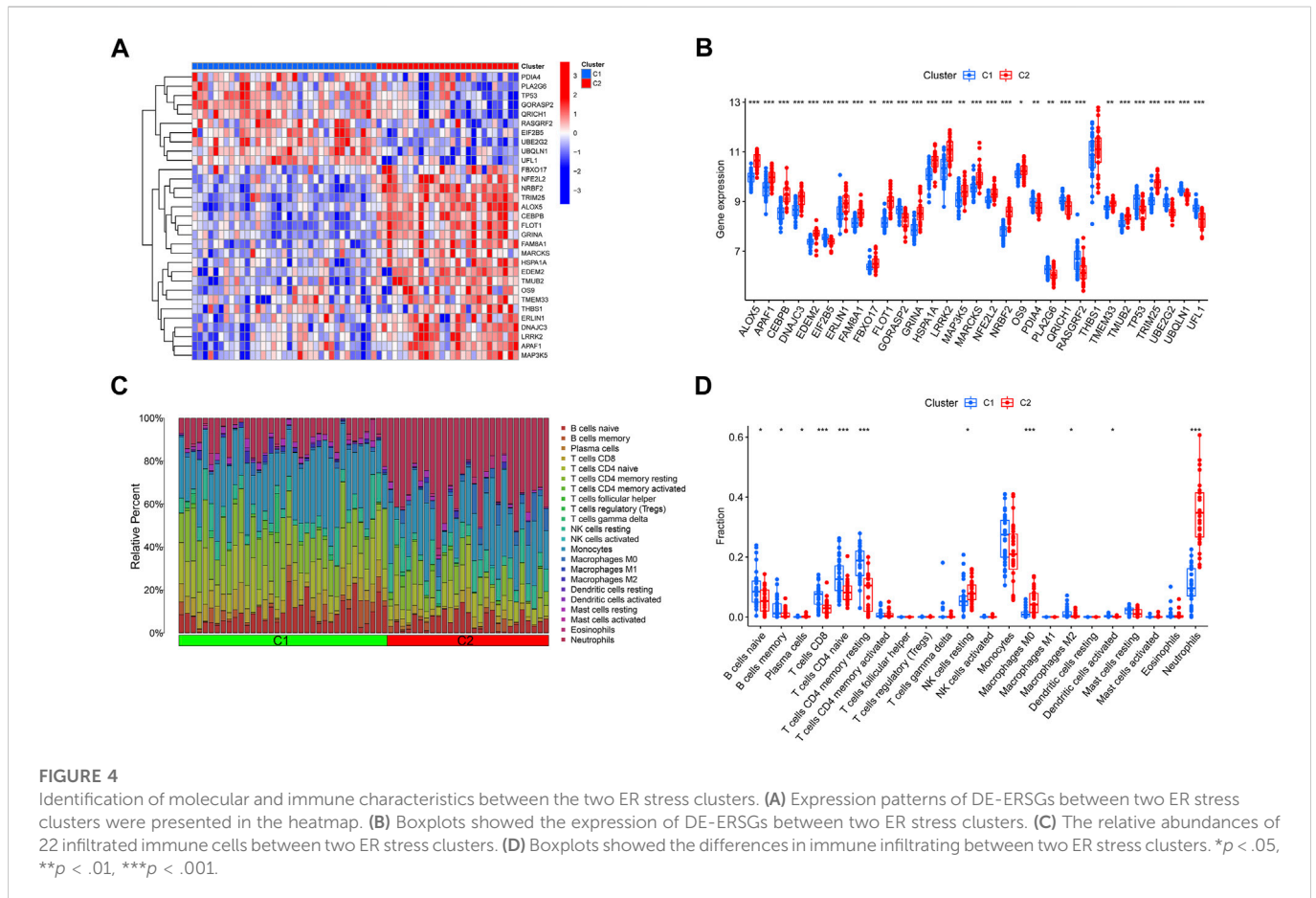
Additionally, only when  $k = 2$ , each subtype's consistency score was  $>.9$  (Figure 3D). Hence, two clusters were finally identified as optimal. Further principal component analysis (PCA) (Ben and Ben, 2021) findings showed a remarkable variation between the aforementioned clusters (Figure 3E).

## Differentiation of ERSGs and immune infiltration features between ERSGs clusters

The expression variations of ERSGs between clusters 1 and 2 were first thoroughly evaluated, allowing for the investigation of the molecular variations across the two categories. Two different ERS patterns were found, each with its unique expression landscape of ERSGs (Figures 4A, B). Furthermore, immune infiltration study findings revealed a changed immune milieu between ER stress clusters 1 and 2 (Figure 4C). Cluster two exhibited higher proportions of macrophages M0, neutrophils, and NK cells resting, whereas the levels of B cells naive, B cells memory, T cells CD8, T cells CD4 memory resting, and monocytes were relatively greater in cluster 1 (Figure 4D). These results confirmed that the two clusters based on the ERSGs had different immune environments, which provided a strong case for foundational genetic changes that link ERS, immune cell infiltration, and the remodeling seen with BPD.

## Screening of gene modules and creation of co-expression networks

To determine which gene modules played a key role in BPD, we used the WGCNA method to create a co-expression network and modules for both



BPD and normal samples. We determined the variation of expression for each gene in GSE32472 and afterward analyzed the top 25% of genes with the greatest variance. Once the soft power value was adjusted to 19, co-expressed gene modules were determined (Figure 5A). By means of the dynamic cutting method, we obtained 14 unique co-expression modules, each of which was assigned a unique color (Figures 5B–D). Module-clinical characteristics (Control and BPD) co-expression was then continually applied to these genes in the 14 colored modules to determine adjacency and similarity. Lastly, 568 genes in the blue module showed the highest correlation with BPD (Figure 5E). In addition, we found a positive link between the blue module and module-related genes (Figure 5F).

We used the WGCNA technique to further scrutinize the key gene modules that are linked to ER stress clusters. To develop a scale-free network, we found that a value of  $\beta = 19$  was the optimal soft threshold (Figure 6A). In particular, a heatmap depicted the TOM of all genes associated with the 13 modules deemed to be significant (Figures 6B–D). An examination of the links between modules and clinical variables (clusters 1 and 2) revealed a strong association between the turquoise module and BPD subtypes (Figure 6E). A correlation study revealed a strong association between turquoise module genes and the chosen module (Figure 6F).

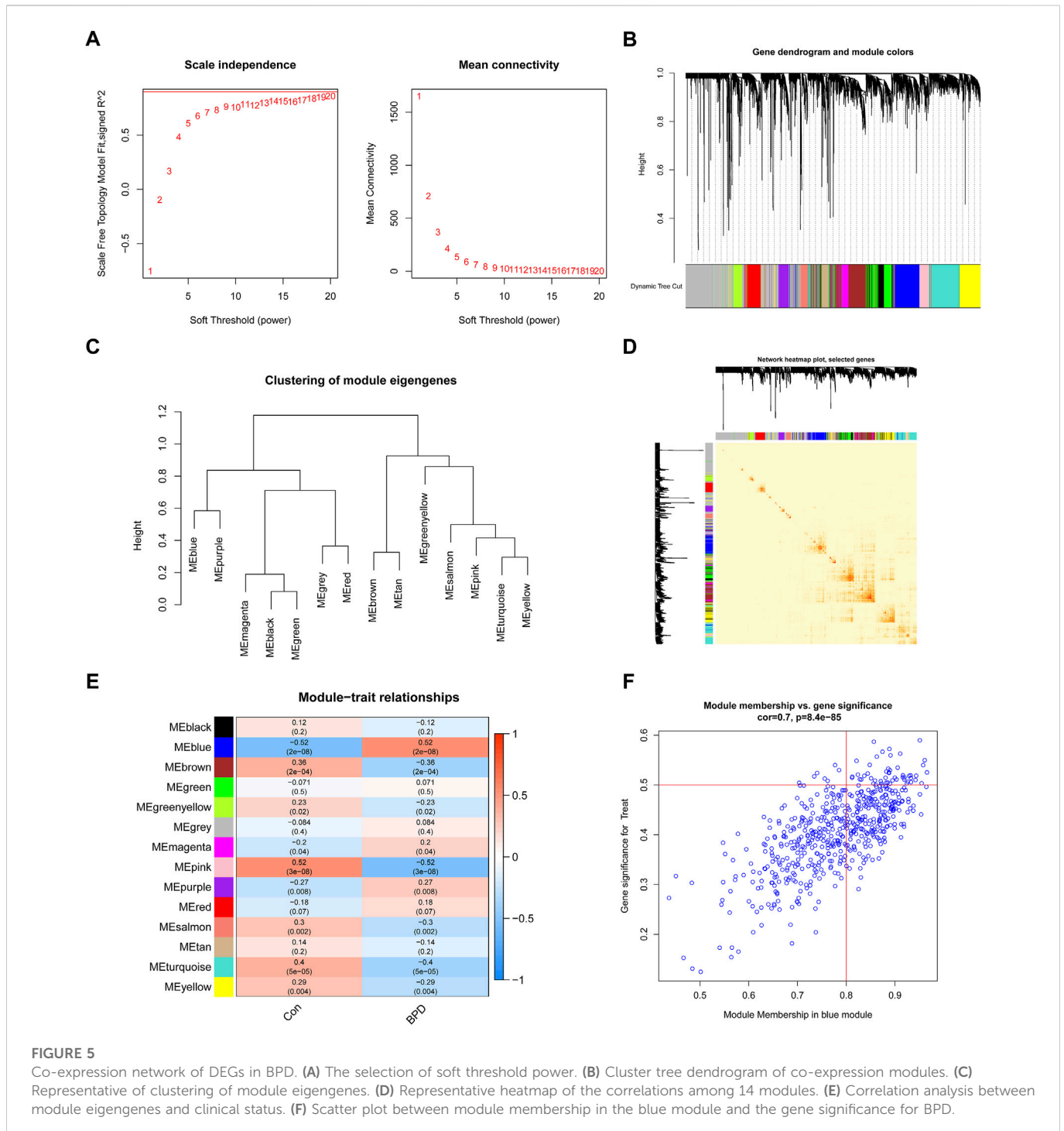
## Determination of DEGs unique to clusters and annotation of their functions

By intersecting the genes involved in ER stress modules with those involved in BPD and controls, we detected 49 cluster-specific DEGs

(Figure 7A). To delve even further into the functional differences between the two clusters, we employed the GSVA analysis to examine the DEGs that were specific to each cluster. The findings illustrated that the regulation of transcription by RNA polymerase 1, regulation of cytoplasmic translational initiation, transcription by RNA polymerase 1, spliceosome, N glycan biosynthesis, aminoacyl tRNA biosynthesis, MYC targets, and MAPK unfolded protein response were enriched in cluster 2, while the icosanoid binding, toll-like receptor 4 binding, respiratory burst, MTOR signaling pathway, long term potentiation, IL6/JAK/STAT3 signaling, inflammatory response, and hypoxia were enriched in cluster 1 (Figures 7B–D).

## Development and evaluation of models for machine learning

We applied four validated machine learning models (SVM, RF, GLM, XGB) using the expression patterns of 49 cluster-specific DEGs in the BPD training dataset to additionally uncover subtype-specific genes with excellent diagnostic significance. The residual variance in SVM and RF machine learning models was quite low (Figures 8A, B). Next, root mean square error (RMSE) was used to determine the rank of importance for each model's top 10 feature variables (Figure 8C). In addition, we computed ROC curves by means of 5-fold cross-validation to assess the discriminative capacity of the four machine learning techniques in the testing dataset. The area under the ROC (AUC) was greatest for the SVM machine-learning model



(AUC) (RF, AUC = .783; SVM, AUC = .798; XGB, AUC = .758; GLM, AUC = .679) (Figure 8D). Altogether, these findings show that the SVM model is superior at distinguishing across patient populations. In the end, the variables PYGL, YIPF1, SLC2A14, CKAP4, and PDLIM7 that were shown to be the most significant by the SVM model were chosen to serve as predictor genes for the subsequent research.

To conduct a more in-depth analysis of the prediction power of the SVM model, we initially designed a nomogram to predict the likelihood of ER stress clusters occurring in 62 BPD patients (Figure 9A). Both the

DCA and calibration curves were used to evaluate the nomogram model's capacity for accurate prediction. The calibration curve indicated that the variation between the observed and anticipated risks of BPD clusters was low (Figure 9B), and DCA shows that our nomogram is quite accurate, highlighting that it might be used to guide medical judgment (Figure 9C). Next, we used GSE108756 to evaluate our 5-gene prognostic model, and its ROC curves demonstrated strong performance (AUC = 1.000) (Figure 9D), demonstrating that our diagnostic model is accurate in differentiating BPD cases from non-BPD individuals.

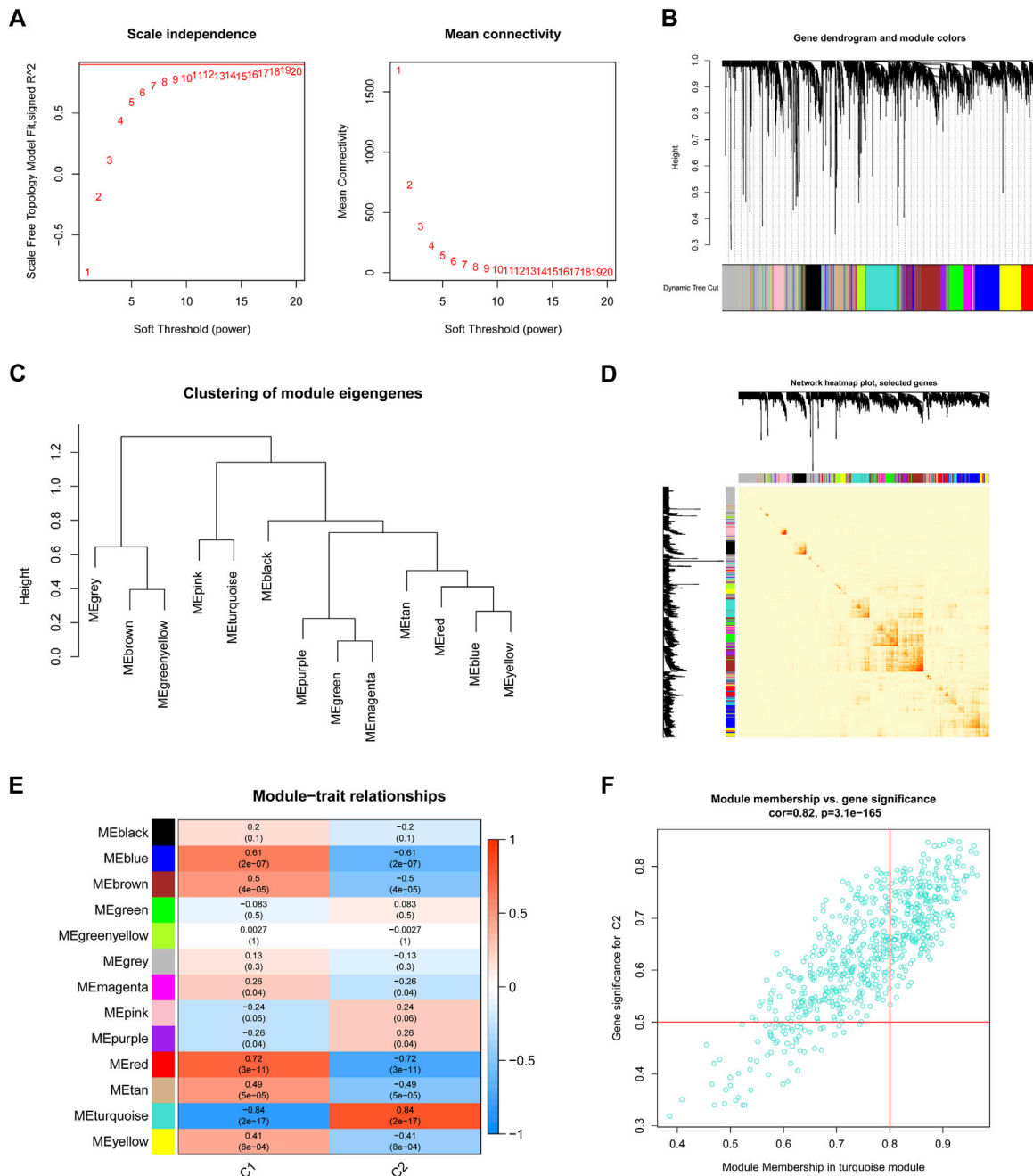


FIGURE 6

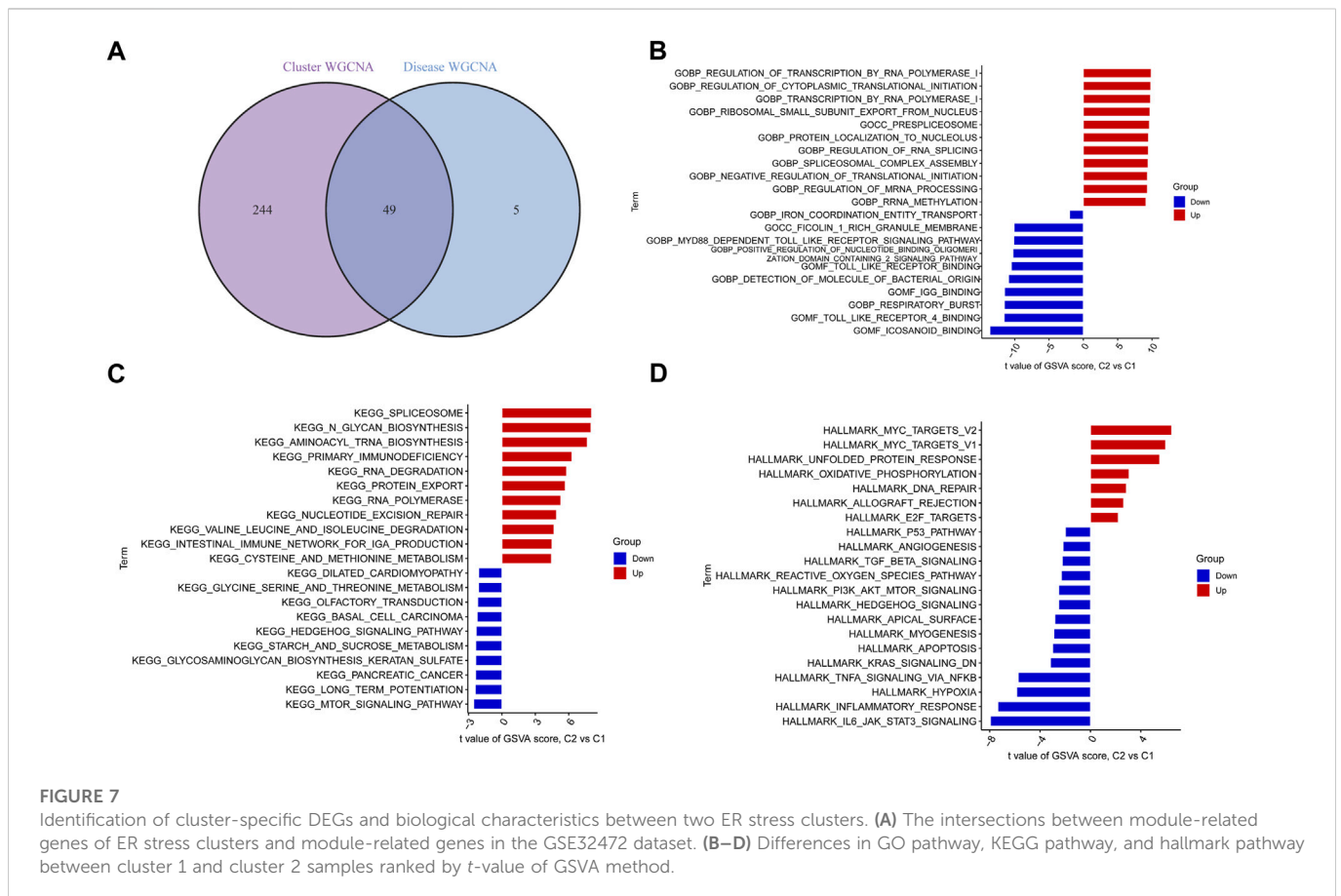
Co-expression network of DEGs between the two ER stress clusters. (A) The selection of soft threshold power. (B) Cluster tree dendrogram of co-expression modules. (C) Representative of clustering of module eigengenes. (D) Representative heatmap of the correlations among 13 modules. (E) Correlation analysis between module eigengenes and clinical status. (F) Scatter plot between module membership in turquoise module and the gene significance for cluster 2.

## Discussion

BPD is a complicated illness that has a significant genetic component. Some scholars contend that a variety of genetic pathways and mutations might be linked to the predisposition to developing BPD (Higano et al., 2021). Each genetic variation may participate in the development of the illness, and the accumulation of these variants results in the dysregulation of biological processes in the growing lungs of preterm babies

(Jung et al., 2019). This high-risk population will benefit from the development of new treatments once the processes behind BPD are better understood. Better molecular clusters must be identified to provide direction for the customized therapy of BPD patients because of the heterogeneity of BPD pathology. The ER is the subcellular organelle that offers a one-of-a-kind environment for the synthesis of cholesterols, lipids, and proteins, as well as for the metabolism of carbohydrates and the storage of calcium (Parmar and Schroder, 2012; Sanvictores and Davis, 2022). The ER is



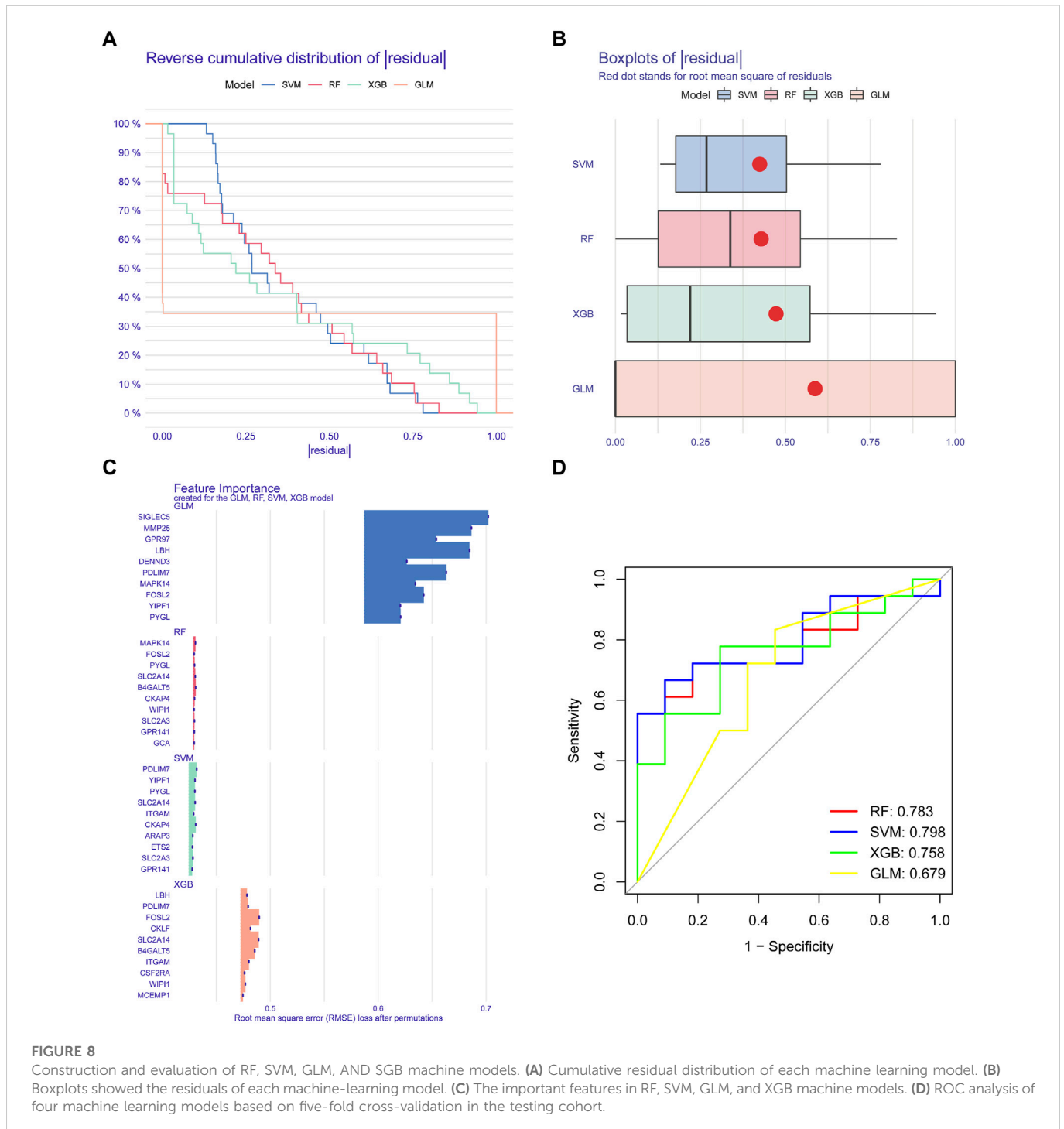


essential to the process of alveoli production in the following ways: 1) preserving the function of the mitochondria; 2) correction of growth factor and receptor post-translation modifications; 3) reducing oxidative stress production; and 4) control of inflammation (Chan et al., 2016; Biber and Isakson, 2017; Koksall et al., 2021). It has been observed that HOX and interferon-gamma increase ER stress in the lungs of newborn mice, thus causing lung damage and the arrival of neutrophils (Ileriturk et al., 2022). Nonetheless, further research is necessary to confirm the exact mechanisms underpinning ER stress and its modulatory involvement in different disorders. As a result, we aimed to better understand the function of ERSGs in the characterization of BPD phenotypes and the immune milieu in which they exist. Furthermore, ER stress-related gene signatures were used to classify BPD cases.

In this work, for the first time, we executed an in-depth comparison of the expression patterns of ERSGs in BPD samples with those from control groups. Compared to non-BPD patients, those with BPD were more likely to have dysregulated ERSGs, highlighting the important function of ERSGs in the development of BPD. We next determined the inter-ERSG connection to fully understand the link between ERSGs and BPD. We found that the presence of ERSG interplay in BPD patients offered evidence that some ERSGs exhibited strong synergistic or antagonistic impacts. Changes in immune cell abundance between control and BPD samples are indicative of the potential involvement of immune cells in the pathogenesis of BPD. In addition, we employed unsupervised cluster analysis to

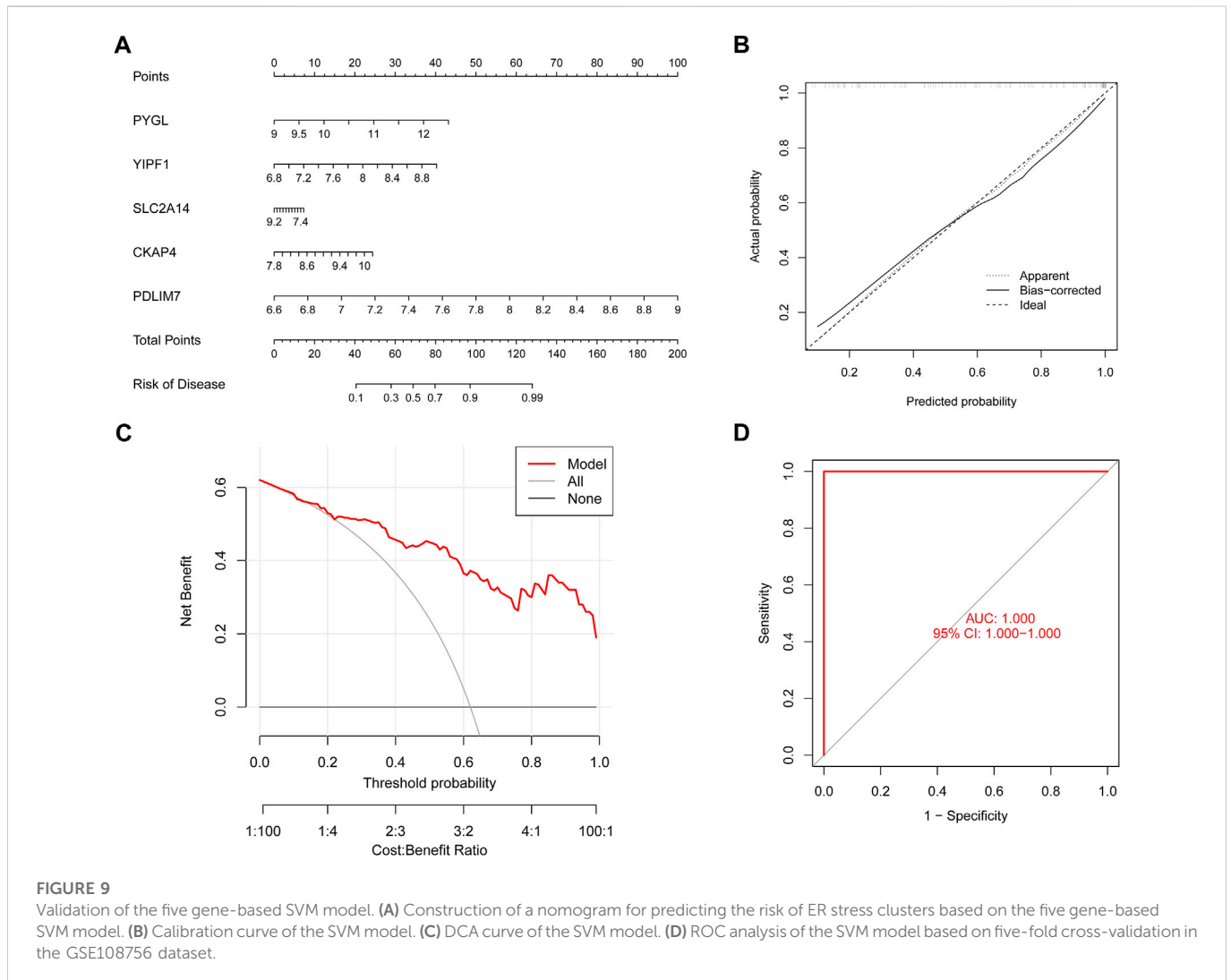
depict various ER stress trends in BPD patients according to the expression patterns of ERSGs, and we found that there were two different ER stress-related clusters. Our results found that Cluster 2 had significantly higher neutrophils. Neutrophils play a very important role in acute lung inflammation in mature and developing organisms (Beck-Schimmer et al., 2005; Phillipson and Kubes, 2011; Aggarwal et al., 2014). During the initiation of inflammation, neutrophils undergo many changes in gene expression and functional properties (Christensen, 1989). Neutrophils originate in the bone marrow and are then released into the circulatory system where they act as the first line of cellular immune defense when they are recruited to the site of injury (Furze and Rankin, 2008; Prame et al., 2018). After the first wave of inflammation, subsequent stages recruit monocytes. Neutrophils remove pathogens through phagocytosis and by releasing proteases, reactive oxygen species (ROS) and bioactive membrane vesicles (Nauseef and Borregaard, 2014). By classifying patients with BPD into Cluster 1 and Cluster 2 subtypes, we have shown that the two types of immune cells have differences, suggesting that there may be differences in outcomes that need to be explored further in the future.

Recent years have seen a surge in the use of machine learning algorithms that incorporate demographic and imaging parameters to anticipate BPD prevalence, and evidence from these research reports shows that multifactorial analyses properly accounted for links among variables, as a consequence, having a reduced error rate and more robust outcomes than univariate analysis. Premised on the expression patterns of cluster-specific DEGs, we generated



an SVM-based prediction model and contrasted the prediction accuracy across four chosen machine learning classifiers (XGB, GLM, SVM, and RF), which demonstrated the best predictive efficiency in the validation cohort, revealing that SVM-based machine learning performs well in classifying BPD subtypes. Then, we built a five-gene SVM model by selecting PYGL, YIPF1, SLC2A14, CKAP4, and PDLIM7 as the most crucial variables. There is evidence that points to PYGL being involved in glycogen degradation (Zhan et al., 2021). The hypoxia metabolism gene PYGL was discovered to be upregulated in many malignancies, including breast cancers and head and neck

squamous cell carcinomas (Kawakubo-Yasukochi et al., 2021; Zhao et al., 2021). The glycogen degradation mediated by PYGL has been hypothesized to sustain the growth of cancer cells. Recently, Wang et al. (2022) reported that PYGL is a potential target of miR-155-5p for regulating the function of pulmonary artery smooth muscle cells in response to hypoxia. YIPF1 belongs to the YIPF family, and HA-tagged YIPF1 has also been observed to partially localize to the ER (Soonthornsit et al., 2017). The SLC2A14 gene is responsible for encoding the glucose transporter member 14 (GLUT14). Disorders of the central nervous system, rheumatoid arthritis, lymphoma, and intraocular pressure in primary open-angle glaucoma have all



been linked to mutations in the SLC2A14 gene (Amir et al., 2016; Amir et al., 2017). CKAP4 is an ER protein that is present on the surface of cell membranes and has been identified to be a cell surface receptor for many proteins such as t-PA and APF (Kimura et al., 2016; Osugi et al., 2019; Harada et al., 2020). Interaction between PDLIM7 and synaptopodin was discovered to occur through several domains, and PDLIM7 colocalizes with synaptopodin on the cisternal organelle, an unusual stacking of ER cisterns that resembles the spine apparatus, and is found at axon initial segments of a subset of neurons (D’Cruz et al., 2016).

This additional external validation in GSE108756 dataset demonstrates that the five gene-based SVM model could reliably predict BPD, expanding our understanding of the diagnostic process for BPD. Moreover, we used PYGL, YIPF1, SLC2A14, CKAP4, and PDLIM7 to develop a nomogram model for diagnosing BPD subtypes. In our tests, the model showed considerable predictive performance, suggesting its potential use in clinical settings. Altogether, the SVM model integrating these five genes provides sufficient guidance for distinguishing between BPD subtypes.

There are a few drawbacks to this research that must be mentioned. To begin, the present research relied heavily on

bioinformatics, and further clinical or experimental evaluation is needed to confirm the expression patterns of ERSGs, which was not the case in our investigation. Additionally, the performance of the prediction model needs to be verified by a more comprehensive set of clinical parameters. Further investigation into the possible relationship between ERSGs and immunological responses is also warranted, and a larger number of BPD samples is required to confirm the reliability of ER stress-related clusters. Even though we used an external dataset to ensure accuracy, further trials are needed to establish a link between the characteristics of BPD pathology.

## Conclusion

Overall, our research revealed the connection between ERSGs and infiltrated immune cells and shed light on the substantial heterogeneity of the immune across BPD patients with diverse ER stress clusters. The best machine learning model for determining BPD subtypes and the clinical prognosis of BPD patients was found to be a five-gene-based SVM model. Our findings clarify the molecular processes driving BPD

heterogeneity and establish for the first time the significance of ER stress in this disorder.

## Data availability statement

All of the research data used in this study may be found in GEO. This database is the source for all multiple microarrays, including GSE32472 and GSE108756. R code and raw data have been uploaded to <https://www.jianguoyun.com/p/DfJaN6sQk4G1CRiO4eIEIAA>.

## Author contributions

ZT and YM collected the data and drafted the manuscript. YH, CW, and XT participated in the analysis and writing of the manuscript. JW and NZ contributed to the study by facilitating insightful discussions. YB, FJ, and FL assisted in generating the concept for the research. The manuscript was written collaboratively, and all authors reviewed and consented to the final manuscript.

## References

- Adamopoulos, C., Farmaki, E., Spilioti, E., Kiaris, H., Piperi, C., and Papavassiliou, A. G. (2014). Advanced glycation end-products induce endoplasmic reticulum stress in human aortic endothelial cells. *Clin. Chem. Lab. Med.* 52, 151–160. doi:10.1515/cclm-2012-0826
- Aggarwal, N. R., King, L. S., and D'Alessio, F. R. (2014). Diverse macrophage populations mediate acute lung inflammation and resolution. *Am. J. Physiol. Lung Cell Mol. Physiol.* 306, L709–L725. doi:10.1152/ajplung.00341.2013
- Amir, S. M., Murphy, B., and Eck, P. (2016). The SLC2A14 gene: Genomic locus, tissue expression, splice variants, and subcellular localization of the protein. *Biochem. Cell Biol.* 94, 331–335. doi:10.1139/bcb-2015-0089
- Amir, S. M., Zhouyao, H., Tu, H., El-Gabalawy, H., Crow, G. H., Levine, M., et al. (2017). The SLC2A14 gene, encoding the novel glucose/dehydroascorbate transporter GLUT14, is associated with inflammatory bowel disease. *Am. J. Clin. Nutr.* 106, 1508–1513. doi:10.3945/ajcn.116.147603
- Bancalari, E., and Jain, D. (2019). Bronchopulmonary dysplasia: 50 Years after the original description. *Neonatology*, 384–391. doi:10.1159/000497422
- Beck-Schimmer, B., Schwendener, R., Pasch, T., Reyes, L., Booy, C., and Schimmer, R. C. (2005). Alveolar macrophages regulate neutrophil recruitment in endotoxin-induced lung injury. *Respir. Res.* 6, 61. doi:10.1186/1465-9921-6-61
- Ben, S. K., and Ben, A. A. (2021). Principal component analysis (PCA). *Tunis. Med.* 2021, 383–389.
- Biwer, L. A., and Isakson, B. E. (2017). Endoplasmic reticulum-mediated signalling in cellular microdomains. *Acta Physiol. (Oxf)* 219, 162–175. doi:10.1111/apha.12675
- Bonadies, L., Zaramella, P., Porzionato, A., Perilongo, G., Muraca, M., and Baraldi, E. (2020). Present and future of bronchopulmonary dysplasia. *J. Clin. Med.* 9, 1539. doi:10.3390/jcm9051539
- Boso, F., Broyda, S. V., and Tartakovsky, D. M. (2014). Cumulative distribution function solutions of advection-reaction equations with uncertain parameters. *Proc. Math. Phys. Eng. Sci.* 470, 20140189. doi:10.1098/rspa.2014.0189
- Chan, P., Stolz, J., Kohl, S., Chiang, W. C., and Lin, J. H. (2016). Endoplasmic reticulum stress in human photoreceptor diseases. *Brain Res.* 1648, 538–541. doi:10.1016/j.brainres.2016.04.021
- Chen, A. C., Burr, L., and McGuckin, M. A. (2018). Oxidative and endoplasmic reticulum stress in respiratory disease. *Clin. Transl. Immunol.* 7, e1019. doi:10.1002/cti2.1019
- Christensen, R. D. (1989). Hematopoiesis in the fetus and neonate. *Pediatr. Res.* 26, 531–535. doi:10.1203/00006450-198912000-00001
- D'Cruz, R., Plant, P. J., Pablo, L. A., Lin, S., Chackowicz, J., Correa, J., et al. (2016). PDLIM7 is a novel target of the ubiquitin ligase Nedd4-1 in skeletal muscle. *Biochem. J.* 473, 267–276. doi:10.1042/BJ20150222
- Furze, R. C., and Rankin, S. M. (2008). Neutrophil mobilization and clearance in the bone marrow. *Immunology* 125, 281–288. doi:10.1111/j.1365-2567.2008.02950.x
- Gilfillan, M., Bhandari, A., and Bhandari, V. (2021). Diagnosis and management of bronchopulmonary dysplasia. *BMJ* 2021, n1974. doi:10.1136/bmj.n1974
- Giraldo, B., Garde, A., Arizmendi, C., Jane, R., Benito, S., Diaz, I., et al. (2006). Support vector machine classification applied on weaning trials patients. *Conf. Proc. IEEE Eng. Med. Biol. Soc.* 2006, 5587–5590. doi:10.1109/IEMBS.2006.259440
- Hanzelmann, S., Castelo, R., and Guinney, J. (2013). Gsva: Gene set variation analysis for microarray and RNA-seq data. *BMC Bioinforma.* 14, 7. doi:10.1186/1471-2105-14-7
- Harada, T., Sada, R., Osugi, Y., Matsumoto, S., Matsuda, T., Hayashi-Nishino, M., et al. (2020). Palmitoylated CKAP4 regulates mitochondrial functions through an interaction with VDAC2 at ER-mitochondria contact sites. *J. Cell Sci.* 133, jcs249045. doi:10.1242/jcs.249045
- Higano, N. S., Ruoss, J. L., and Woods, J. C. (2021). Modern pulmonary imaging of bronchopulmonary dysplasia. *J. Perinatol.* 41, 707–717. doi:10.1038/s41372-021-00929-7
- Hu, Y., Yang, W., Xie, L., Liu, T., Liu, H., and Liu, B. (2020). Endoplasmic reticulum stress and pulmonary hypertension. *Pulm. Circ.*, 766647591. doi:10.1177/2045894019900121
- Ileriturk, M., Kandemir, O., and Kandemir, F. M. (2022). Evaluation of protective effects of quercetin against cypermethrin-induced lung toxicity in rats via oxidative stress, inflammation, apoptosis, autophagy, and endoplasmic reticulum stress pathway. *Environ. Toxicol.* 37 (2022), 2639–2650. doi:10.1002/tox.23624
- Jung, Y. H., Choi, C. W., and Kim, B. I. (2019). Strategies to improve outcomes of bronchopulmonary dysplasia. *Korean J. Pediatr.* 2019, 380–381. doi:10.3345/kjp.2019.00661
- Kawakubo-Yasukochi, T., Yano, E., Kimura, S., Nishinakagawa, T., Mizokami, A., Hayashi, Y., et al. (2021). Hepatic glycogenolysis is determined by maternal high-calorie diet via methylation of Pygl and it is modified by oteocalcin administration in mice. *Mol. Metab.* 2021, 101360. doi:10.1016/j.molmet.2021.101360
- Kepp, O., and Galluzzi, L. (2020). Preface: Endoplasmic reticulum in health and disease. *Int. Rev. Cell Mol. Biol.* 350, xiii. doi:10.1016/S1937-6448(20)30020-4
- Kimura, H., Fumoto, K., Shojima, K., Nojima, S., Osugi, Y., Tomihara, H., et al. (2016). CKAP4 is a Dickkopf1 receptor and is involved in tumor progression. *J. Clin. Invest.* 126, 2689–2705. doi:10.1172/JCI84658
- Koksal, A. R., Verne, G. N., and Zhou, Q. (2021). Endoplasmic reticulum stress in biological processing and disease. *J. Invest. Med.* 69 (2021), 309–315. doi:10.1136/jim-2020-001570
- Langfelder, P., and Horvath, S. (2008). Wgcna: an R package for weighted correlation network analysis. *BMC Bioinforma.* 9, 559. doi:10.1186/1471-2105-9-559
- Lee, Y., and Nelder, J. A. (2002). Analysis of ulcer data using hierarchical generalized linear models. *Stat. Med.* 21, 191–202. doi:10.1002/sim.978
- Leprivier, G., Rotblat, B., Khan, D., Jan, E., and Sorensen, P. H. (2015). Stress-mediated translational control in cancer cells. *Biochim. Biophys. Acta* 1849, 845–860. doi:10.1016/j.bbagr.2014.11.002
- Marciniak, S. J. (2017). Endoplasmic reticulum stress in lung disease. *Eur. Respir. Rev.* 26, 170018. doi:10.1183/16000617.0018-2017
- Nauseef, W. M., and Borregaard, N. (2014). Neutrophils at work. *Nat. Immunol.* 15, 602–611. doi:10.1038/ni.2921

## Conflict of interest

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

## Publisher's note

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

## Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fphys.2022.1084650/full#supplementary-material>

- Newman, A. M., Liu, C. L., Green, M. R., Gentles, A. J., Feng, W., Xu, Y., et al. (2015). Robust enumeration of cell subsets from tissue expression profiles. *Nat. Methods* 12, 453–457. doi:10.1038/nmeth.3337
- Osugi, Y., Fumoto, K., and Kikuchi, A. (2019). CKAP4 regulates cell migration via the interaction with and recycling of integrin. *Mol. Cell. Biol.* 39, 000733–19. doi:10.1128/MCB.00073-19
- Parmar, V. M., and Schroder, M. (2012). Sensing endoplasmic reticulum stress. *Adv. Exp. Med. Biol.* 738, 153–168. doi:10.1007/978-1-4614-1680-7\_10
- Phillipson, M., and Kubes, P. (2011). The neutrophil in vascular inflammation. *Nat. Med.* 17, 1381–1390. doi:10.1038/nm.2514
- Prame, K. K., Nicholls, A. J., and Wong, C. (2018). Partners in crime: Neutrophils and monocytes/macrophages in inflammation and disease. *Cell Tissue Res.* 2018, 551–565. doi:10.1007/s00441-017-2753-2
- Pritchard, K. A., Jing, X., Teng, M., Wells, C., Jia, S., Afolayan, A. J., et al. (2022). Role of endoplasmic reticulum stress in impaired neonatal lung growth and bronchopulmonary dysplasia. *PLoS One* 17, e0269564. doi:10.1371/journal.pone.0269564
- Rigatti, S. J. (2017). Random forest. *J. Insur Med.* 47, 31–39. doi:10.17849/inm-47-01-31-39.1
- Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J. C., et al. (2011). pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinforma.* 12, 77. doi:10.1186/1471-2105-12-77
- Sanvictores, T., and Davis, D. D. (2022). *Histology, rough endoplasmic reticulum*. Treasure Island: Statpearls.
- Shukla, V. V., and Ambalavanan, N. (2021). Recent advances in bronchopulmonary dysplasia. *Indian J. Pediatr.* 88, 690–695. doi:10.1007/s12098-021-03766-w
- Sillers, L., Alexiou, S., and Jensen, E. A. (2020). Lifelong pulmonary sequelae of bronchopulmonary dysplasia. *Curr. Opin. Pediatr.* 32, 252–260. doi:10.1097/MOP.0000000000000884
- Soonthornsit, J., Sakai, N., Sasaki, Y., Watanabe, R., Osako, S., and Nakamura, N. (2017). YIPF1, YIPF2, and YIPF6 are medial-/trans-Golgi and trans-Golgi network-localized Yip domain family proteins, which play a role in the Golgi reassembly and glycan synthesis. *Exp. Cell Res.* 353, 100–108. doi:10.1016/j.yexcr.2017.03.011
- Thebaud, B., Goss, K. N., Laughon, M., Whitsett, J. A., Abman, S. H., Steinhorn, R. H., et al. (2019). Bronchopulmonary dysplasia. *Nat. Rev. Dis. Prim.* 5, 78. doi:10.1038/s41572-019-0127-7
- Vickers, A. J., and Elkin, E. B. (2006). Decision curve analysis: A novel method for evaluating prediction models. *Med. Decis. Mak.* 26, 565–574. doi:10.1177/0272989X06295361
- Wang, G., Tao, X., and Peng, L. (2022). miR-155-5p regulates hypoxia-induced pulmonary artery smooth muscle cell function by targeting PYGL. *Bioengineered* 13 (2022), 12985–12997. doi:10.1080/21655979.2022.2079304
- Wilkerson, M. D., and Hayes, D. N. (2010). ConsensusClusterPlus: A class discovery tool with confidence assessments and item tracking. *Bioinformatics* 26, 1572–1573. doi:10.1093/bioinformatics/btq170
- Zeeshan, H. M., Lee, G. H., Kim, H. R., and Chae, H. J. (2016). Endoplasmic reticulum stress and associated ROS. *Int. J. Mol. Sci.* 17, 327. doi:10.3390/ijms17030327
- Zhan, Q., Lv, Z., Tang, Q., Huang, L., Chen, X., Yang, M., et al. (2021). Glycogen storage disease type VI with a novel PYGL mutation: Two case reports and literature review. *Med. Baltim.* 2021, e25520. doi:10.1097/MD.00000000000025520
- Zhao, C. Y., Hua, C. H., Li, C. H., Zheng, R. Z., and Li, X. Y. (2021). High PYGL expression predicts poor prognosis in human gliomas. *Front. Neurol.* 2021, 652931. doi:10.3389/fneur.2021.652931