Integrating Machine Learning and Multiomics Analyses to Identify Immune-Related Biomarkers and Mechanisms in Primary Biliary Cholangitis

Zhiyu Zeng, PhD[1,2,*], Jian Fang, MD[3,*], Li Chen, PhD[1,2], Ying Liang, MD[1], Dongliang Li, PhD[1,2], Lei Xia, MD[3] and Longke Xie, MD[4,5]

INTRODUCTION: **Primary biliary cholangitis (PBC) is a chronic autoimmune liver disease that gradually progresses, making early diagnosis and treatment challenging. Reliable biomarkers could enhance diagnostic accuracy and therapeutic development.**

METHODS: **This study analyzed 3 publicly available gene expression data sets from the Gene Expression Omnibus database: GSE119600 (90 patients with PBC and 47 healthy controls), GSE159676 (12 PBC patients and 6 controls), and GSE61260 (11 patients with PBC and 38 controls). To identify genes closely linked to PBC, we applied machine learning techniques, including Least Absolute Shrinkage and Selection Operator, Support Vector Machine-Recursive Feature Elimination, and random forest. We subsequently conducted gene set enrichment and immune cell infiltration analyses to investigate their biological significance. IN addition, potential drug interactions were explored through the Drug Gene Interaction Database, and a competing endogenous RNA regulatory network was developed to examine gene regulation. Finally, the expression of selected genes was validated through multiplex immunofluorescence staining of liver tissue samples from patients with PBC.**

RESULTS: **We identified proteasome subunit beta 7, TRAF family member associated nuclear factor kappa-light-chain-enhancer of activated B cells activator Albumin (TANK)-binding kinase 1, solute carrier family 29 member 1, and natural killer cell receptor 2B4 as key genes associated with PBC; these genes were significantly enriched in immune-related pathways and strongly correlated with immune regulation. Drug target prediction indicated that some genes could interact with existing immunomodulators or anticancer drugs. Competing endogenous RNA network analysis revealed that TANK-binding kinase 1, solute carrier family 29 member 1, and natural killer cell receptor 2B4 interact with multiple miRNAs and long noncoding RNAs, potentially regulating the immune microenvironment of PBC through noncoding RNA mechanisms. Immunofluorescence staining confirmed that these genes were highly expressed in liver tissues from patients with PBC.**

DISCUSSION: **By integrating machine learning and functional analyses, this study identified 4 genes that may serve as potential biomarkers for PBC. Their involvement in immune regulation suggests possible applications in both diagnosis and therapy. Further studies are necessary to explore their clinical relevance and therapeutic potential.**

KEYWORDS: primary biliary cholangitis; machine learning; multiplex immunofluorescence staining; mechanism

ABBREVIATIONS: AMA, anti-mitochondrial antibodies; AUC, area under the ROC curve; CC, cellular component; CD244, natural killer cell receptor 2B4; ceRNA, competing endogenous RNAs; CIBERSORT, Cell-type Identification by Estimating Relative Subsets of RNA

[1]Department of Hepatobiliary Disease, Fuzong Clinical Medical College of Fujian Medical University, 900th Hospital of PLA Joint Logistic Support Force, Fuzhou, Fujian, China; [2]Department of Hepatobiliary Disease, 900th Hospital of PLA Joint Logistic Support Force, Fuzhou, China; [3]Department of Hepatobiliary Medicine, The Third Affiliated People's Hospital of Fujian University of Traditional Chinese Medicine, Fuzhou, China; [4]Department of Gastroenterology, Fuzong Clinical Medical College of Fujian Medical University, 900th Hospital of PLA Joint Logistic Support Force, Fuzhou, Fujian, China; [5]Department of Gastroenterology, 900th Hospital of PLA Joint Logistic Support Force, Fuzhou, China. **Correspondence:** Lei Xia, MD. E-mail: 258154181@qq.com. Longke Xie, MD. E-mail: 1075072560@qq.com.
*Zhiyu Zeng, Jian Fang contributed equally to this work.
Received May 18, 2025; accepted August 14, 2025; published online August 22, 2025

Transcripts; DEGs, differentially expressed genes; DGIdb, Drug Gene Interaction Database; ENT1, equilibrative nucleoside transporter; GEO, Gene Expression Omnibus; GO, Gene Ontology; GSEA, gene set enrichment analysis; H&E, hematoxylin and eosin; HRP, horseradish peroxidase; KEGG, Kyoto Encyclopedia of Genes and Genomes; LASSO, Least Absolute Shrinkage and Selection operator; lncRNAs, long noncoding RNA; MF, molecular function; MSigDB, Molecular Signatures Database; NCBI, National Center for Biotechnology Information; PBC, primary biliary cholangitis; PBS, phosphate-buffered saline; pROC, receiver operating characteristic; PSMB7, proteasome subunit beta type-7; RF, random forest; RFE, recursive feature elimination; RMSE, root mean square error; ROC, receiver operating characteristic; SLC29A1, solute carrier family 29 member 1; SVM, support vector machine; TBK1, TANK-binding kinase 1; TSA, tyramide signal amplification

SUPPLEMENTARY MATERIAL accompanies this paper at http://links.lww.com/CTG/B365, http://links.lww.com/CTG/B366

Clinical and Translational Gastroenterology 2025;16:e00907. https://doi.org/10.14309/ctg.0000000000000907

INTRODUCTION

Primary biliary cholangitis (PBC) is a progressive autoimmune liver disorder that gradually damages the small bile ducts within the liver. If left untreated, it can progress to cirrhosis, liver failure, or hepatocellular carcinoma (1). The condition is more commonly observed in middle-aged women, although its underlying cause remains uncertain. Research suggests that immune system irregularities, particularly disruptions in T-cell-mediated immune responses, are central to the development and progression of PBC (2). Individuals with PBC often experience fatigue, itching, and jaundice (3), symptoms that typically intensify over time. However, in the early stages, these signs are subtle and nonspecific, leading to frequent misdiagnosis as other liver conditions. Currently, the presence of anti-mitochondrial antibodies serves as a key biomarker for diagnosing PBC (4). Nevertheless, a subset of patients may test negative for anti-mitochondrial antibody, complicating diagnosis and highlighting the need for additional reliable biomarkers. Liver histopathology remains a valuable diagnostic tool capable of detecting characteristic PBC features, including chronic nonsuppurative destructive cholangitis and interlobular bile duct damage (5). Despite its diagnostic value, this method is invasive, entails certain risks, and is impractical for widespread early diagnosis and screening. Therefore, the identification of novel noninvasive biomarkers and improvements in early PBC detection accuracy have emerged as critical research topics.

With the rapid development of genomics, transcriptomics, and high-throughput technologies, significant progress has been made in understanding the molecular mechanisms of PBC. Studies have shown that PBC is closely associated with immune system disorders and involves multiple biological processes, including cytokine activity and bile acid metabolism (6). However, despite existing studies revealing several molecular mechanisms of PBC, comprehensive systemic analyses of the exact pathogenesis, especially the role of immune cells within the immune microenvironment and their interactions with PBC-related genes, are still lacking. Moreover, ursodeoxycholic acid is currently the only first-line treatment of PBC, but up to 40% of patients with PBC exhibit an incomplete response, and its therapeutic efficacy remains limited (7), further highlighting the urgent need to identify new molecular targets and therapeutic strategies.

Therefore, the aim of this study was to apply machine learning algorithms to integrate gene expression data from multiple public data sets to identify potential key genes associated with PBC diagnosis. Through differential gene analysis, gene set enrichment analysis (GSEA), and immune cell infiltration analysis, we explored the biological functions and potential mechanisms of these genes in PBC. On the basis of these analyses, we further collected liver tissue samples from patients with PBC and healthy individuals to validate the expression levels and tissue localization of the identified key genes by immunofluorescence staining experiments. We anticipate that the novel biomarkers identified in this study will increase the accuracy of early PBC diagnosis and provide a theoretical basis for understanding the immune regulatory mechanisms and targeted therapies for PBC, thereby promoting precision diagnosis and treatment of this disease.

MATERIALS AND METHODS

Data acquisition and processing

In this study, PBC data were retrieved from the Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo/) database. Using the keyword "PBC," we downloaded the data sets GSE119600 (47 healthy controls and 90 patients with PBC) (8), GSE159676 (6 healthy controls and 12 patients with PBC) (9), and GSE61260 (38 healthy controls and 11 patients with PBC) (10). After normalization, the data sets were analyzed and visualized using R version 4.2.2 and the R packages limma (v3.52.4), sva (v3.44.0), and ggplot2 (v3.4.0).

Differentially expressed gene analysis and functional enrichment analysis

To identify differentially expressed genes (DEGs) in PBC, we set the screening criteria as $|\log FC| > 0.5$ and $P < 0.05$, using the limma package for DEG detection. Gene Ontology (GO) enrichment analysis and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis were subsequently performed to explore the potential biological functions of these DEGs in PBC. GO and KEGG enrichment analyses were conducted using the R packages ggplot2 (v3.4.0), clusterProfiler (v4.4.4), org.Hs.eg.db (v3.15.0), and enrichplot (v1.16.1), with significant GO and KEGG pathways identified under the condition of adjusted $P < 0.05$.

Screening of key diagnostic genes for PBC

To identify key diagnostic genes for PBC, this study combined 3 machine learning algorithms to screen for PBC-related DEGs. First, the Least Absolute Shrinkage and Selection Operator (LASSO) algorithm and support vector machine (SVM) algorithm were used to classify diagnostic marker genes for PBC, with model parameters optimized through 10-fold cross-validation through the glmnet package (v4.1-4). The SVM-Recursive Feature Elimination (SVM-RFE) algorithm identifies the most discriminative genes by constructing a hyperplane with the

maximum margin in the feature space, implemented through the e1071 (v1.7-13) and svm Radial packages in R. The random forest (RF) algorithm ranks PBC-related DEGs on the basis of the Gini index, identifying characteristic expression genes through recursive binary tree partitioning. To identify genes with the highest classification potential, the RF model was built through the RF package (v4.7-1.1). The intersection of genes identified by all 3 algorithms was considered a key diagnostic gene for PBC. To assess the model's predictive performance, the p receiver operating characteristic package (v1.18.0) was used to compute the area under the receiver operating characteristic curve (AUC). A higher AUC value indicates greater predictive accuracy.

GSEA

GSEA was conducted through the "c2.cp.kegg.v11.0" gene set from the Molecular Signatures Database (MSigDB, http://software.broadinstitute.org/gsea/msigdb) as a reference. This analysis was designed to investigate the biological functions and potential roles of the identified key genes across different signaling pathways. To ensure the robustness of the findings, 1,000 random permutations were performed, and enrichment scores were normalized. Statistical significance was determined on the basis of an adjusted *P* value threshold of less than 0.05.

Immune cell infiltration analysis

To assess the composition of immune cells within tissues, we used the Cell-type Identification by Estimating Relative Subsets Of RNA Transcripts algorithm (http://cibersortx.stanford.edu), which estimates the relative abundance of 22 distinct immune cell subsets on the basis of gene expression data. To further investigate the associations between the expression of key genes and immune cell infiltration, Spearman correlation analysis was performed through R. The resulting correlation patterns were visualized with the ggplot2 package, which provides a clear representation of the interactions between gene expression levels and immune cell infiltration dynamics.

Key gene drug prediction and competing endogenous RNA network construction

To explore potential therapeutic strategies, the Drug Gene Interaction Database (DGIdb) (https://dgidb.org) was used to identify targeted drugs associated with key genes through default screening parameters. In addition, the mRNA-miRNA interactions of key genes were predicted through data from the National Center for Biotechnology Information and the StarBase databases (http://starbase.sysu.edu.cn). On the basis of the prediction results, a competing endogenous RNA (ceRNA) network comprising mRNAs, miRNAs, and long noncoding RNAs (lncRNAs) was constructed. This network was visualized through Cytoscape software (version 3.7.1), revealing the interactions between key genes, miRNAs, and lncRNAs, as well as their potential regulatory mechanisms in PBC.

Multiplex immunofluorescence staining

From January 2024 to January 2025, a total of 11 paraffin-embedded liver tissue samples were obtained from the 900th Hospital of the Joint Logistics Support Force, comprising tissues from one healthy donor and 10 patients diagnosed with PBC. The diagnosis of PBC was determined by clinical experts through the assessment of elevated serum alkaline phosphatase levels, detection of specific autoantibodies, and histopathological examination of liver biopsy samples. Ethical approval for this study was granted by the 900th Hospital's Ethics Committee (Approval No. 20221025), with informed consent obtained from all participants in writing. Liver tissues were fixed with 4% paraformaldehyde, embedded in paraffin and sliced into 4 μm thick sections for subsequent hematoxylin and eosin staining. Multiplex immunofluorescence staining was conducted on paraffin-embedded liver sections from both patients with PBC and healthy controls by a 5-color multiplex fluorescence staining kit (HuBlue Bio, Catalog No. RC0086Plus-45RM). After deparaffinization and rehydration, antigen retrieval of the 4 μm thick liver sections was performed with sodium citrate buffer, and the samples were heated in a microwave at medium power for 8 minutes, followed by an 8-minute rest and additional heating at low-medium power for 7 minutes. To inhibit endogenous peroxidase activity, the sections were treated with 3% hydrogen peroxide, followed by incubation in 5% bovine serum for 30 minutes to ensure effective blocking. The tissue sections were incubated overnight at 4 °C in a dark environment with primary antibodies diluted to specific concentrations: TANK (1:500, Immunoway, YT4540), natural killer cell receptor 2B4 (CD244; 1:200, Immunoway, YT5615), equilibrative nucleoside transporter 1 (1:500, Ab-mart, PC13077s), and proteasome subunit beta 7 (PSMB7) (1:100, Ab-mart, PH1345S). After incubation, the sections were washed with phosphate-buffered saline and incubated with horseradish peroxidase-conjugated secondary antibodies at room temperature for 50 minutes. Tyramide signal amplification fluorescent dyes diluted 1:200 (TYR-520Plus, TYR-570Plus, TYR-620Plus, and TYR-690Plus; HuBlue Bio) were subsequently applied for 7 minutes of staining. After each round of staining, antigen stripping was performed through sodium citrate buffer at 95 °C for 25 minutes in a water bath, and the staining-stripping cycle was repeated 4 times. Finally, the slides were mounted with antifade mounting medium (Meilun Bio, MA0221), and multichannel fluorescence images were captured by a fluorescence microscope (NIKON, ECLIPSE Ci-L) at different magnifications. Quantitative analysis was conducted on the basis of fluorescence intensity and the distribution of positive signals, which were evaluated through image analysis software.

Ethical approval for this study was granted by the 900th Hospital's Ethics Committee (Approval No. 20221025), with informed consent obtained from all participants in writing.

RESULTS

Identification and enrichment analysis of DEGs in PBC

Three data sets (GSE119600, GSE159676, and GSE61260) comprising data from 113 patients with PBC and 91 healthy controls were analyzed in this study. The gene expression matrix for PBC was first processed through normalization (see Supplementary Figure 1A, http://links.lww.com/CTG/B365), and DEGs were subsequently identified through R software. Supplementary Digital Content (see Supplementary Figure 1B, http://links.lww.com/CTG/B365) presents a heatmap illustrating the differential expression of genes between PBC and control samples, revealing 74 DEGs (see Supplementary Table S1, http://links.lww.com/CTG/B365). GO and KEGG pathway enrichment analyses were performed to investigate the functional roles and signaling pathways associated with these DEGs in PBC. According to the GO analysis, the DEGs were predominantly associated with ribosomes, focal adhesions, and cell-matrix junctions in the context of cellular components, whereas molecular functions

were linked mainly to ribosomal structural elements (see Supplementary Table S2, http://links.lww.com/CTG/B365). KEGG pathway analysis revealed that the DEGs were significantly associated with ribosomal functions, motor protein activities, and Toll-like receptor signaling pathways (see Supplementary Table S3, http://links.lww.com/CTG/B365).

Identification of 4 key genes associated with PBC

Considering the marked biological variations between patients with PBC and healthy controls, the aim of this study was to assess the diagnostic value of DEGs in PBC. To identify key genes distinguishing patients with PBC, 3 machine learning techniques—LASSO, SVM-RFE, and RF—were applied to analyze DEGs within the control data set. The LASSO algorithm selected 18 genes from the 74 DEGs (Figure 1a,b; see Supplementary Table S4, http://links.lww.com/CTG/B365). Through the SVM-RFE algorithm, 30 genes were selected, achieving a maximum accuracy of 0.794 and a minimum root mean square error of 0.206 (Figure 1c,d; see Supplementary Table S4, http://links.lww.com/CTG/B365). The RF algorithm evaluated the importance of NAFRGs, identifying 11 genes with importance scores

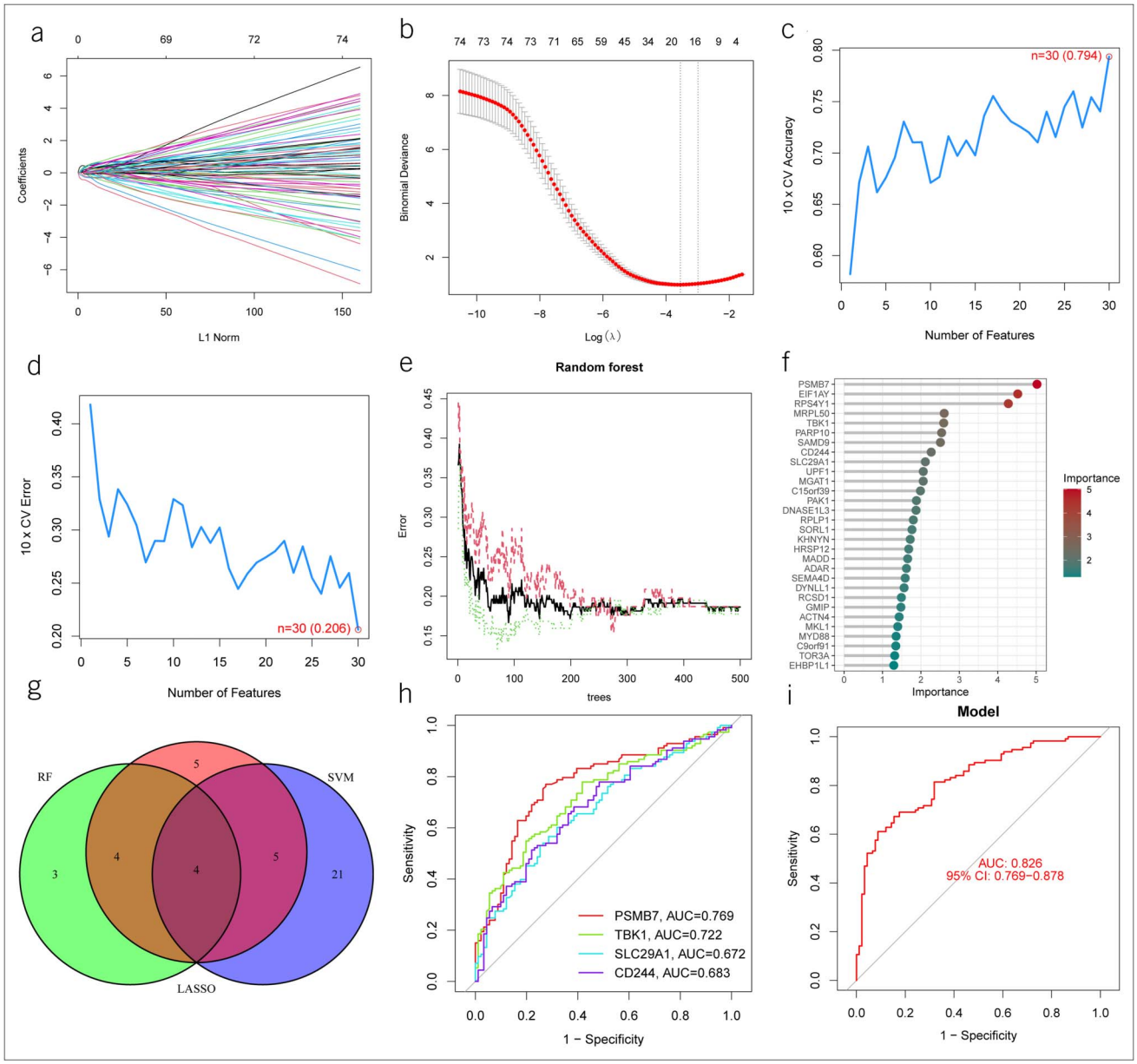

Figure 1. Four key genes were identified as critical for primary biliary cholangitis (PBC). (**a, b**) Least Absolute Shrinkage and Selection Operator (LASSO) logistic regression was used for penalty parameter tuning and tenfold cross-validation, selecting 18 PBC-related features. (**c, d**) The Support Vector Machine-Recursive Feature Elimination (SVM–RFE) algorithm was applied to identify PBC-related genes and determine the optimal combination of key genes, ultimately resulting in the selection of 30 key genes (maximum accuracy = 0.794, minimum root mean square error = 0.206). (**e–f**) The random forest (RF) algorithm was used for key gene selection, identifying 11 genes with importance scores greater than 0.2. (**g**) Intersection of key genes obtained from the LASSO, SVM-RFE, and RF models. (**h**) Receiver operating characteristic (ROC) curves for the 4 key genes in PBC. (**i**) ROC curve of the 4 genes as a combined model for PBC classification.

exceeding 2 (Figure 1e,f; see Supplementary Table S4, http://links.lww.com/CTG/B365). By combining the outputs from the 3 machine learning models, 4 critical genes—PSMB7, TANK-binding kinase 1 (TBK1), solute carrier family 29 member 1 (SLC29A1), and CD244—were identified as significant markers (Figure 1g). Receiver operating characteristic curves were generated to assess the effectiveness of these 4 genes in differentiating PBC samples from healthy controls (Figure 1f). The analysis revealed that these 4 genes had an AUC of 0.826, highlighting their strong diagnostic performance with high accuracy and specificity in distinguishing patients with PBC from healthy controls (Figure 1i). These results imply that the identified genes exhibit robust diagnostic accuracy and specificity in differentiating patients with PBC from healthy individuals.

GSEA analysis of key genes

Given the unclear potential functions of key genes in PBC, GSEA was used to investigate their biological roles and significance in PBC. The GSEA results revealed that PSMB7, TBK1, and SLC29A1 were significantly enriched in multiple immune-related pathways, including those related to the regulation of the actin cytoskeleton, focal adhesion, oxidative phosphorylation, and the phosphatidylinositol signaling system (Figure 2). The enrichment of TBK1 in the chemokine signaling pathway suggests its potential role in immune cell signal transduction.

Immune cell infiltration analysis

To compare the immune microenvironment differences between patients with PBC and healthy controls, the Cell-type Identification by Estimating Relative Subsets Of RNA Transcripts algorithm was used to analyze immune cell infiltration. The immune cell infiltration profiles of the normal and PBC groups are first illustrated (see Supplementary Figure 2A, http://links.lww.com/CTG/B365), followed by a correlation heatmap of 22 immune cell types across both groups (see Supplementary Figure 2B, http://links.lww.com/CTG/B365). The relationships between the 4 core genes and immune genes were then analyzed, revealing strong correlations with multiple immune cell types (see Supplementary Figures 2C–2F, http://links.lww.com/CTG/B365).

Prediction of targeted drugs for key genes and construction of the ceRNA network

The DGIdb database facilitated the identification of potential drugs that target key genes. Cytoscape analysis revealed strong interactions between the PSMB7, SLC29A1, and CD244 genes and approved drugs, primarily anticancer drugs. TBK1 is involved in numerous drug interactions, although most drugs have not yet received approval (see Supplementary Figure 3, http://links.lww.com/CTG/B365). To further explore the regulatory mechanisms of the core genes in PBC, a ceRNA regulatory network was constructed (see Supplementary Figure 4, http://links.lww.com/CTG/B365). TBK1 forms a complex regulatory network with lncRNAs such as RP11-343D2.11 and LINC01070 through miRNAs, including miR-302a-5p and miR-365a-3p. CD244 interacts with miR-876-5p, miR-26b-3p, and lncRNAs such as RP11-699C17.1 and RP13-143G15.4. SLC29A1 is regulated by miRNAs such as miR-185-5p and miR-149-3p, along with lncRNAs such as LINC00266 and LINC00666.

Validation of key gene expression in clinical samples

To further validate the expression levels and spatial distribution of PSMB7, TBK1, SLC29A1, and CD244 in the liver tissues of patients with PBC, liver tissue samples from one healthy individual and 10 patients with PBC were analyzed by multiplex immunofluorescence staining. The results revealed significantly increased expression levels of PSMB7, TANK, equilibrative nucleoside transporter 1, and CD244 antibodies in the liver tissues of patients with PBC compared with those in normal liver tissues. Notably, high expression of these key gene-related proteins was observed in both inflammatory cells and hepatocytes, particularly in the portal areas of patients with PBC (Figures 3–6, see Supplementary Figure 5, http://links.lww.com/CTG/B365).

DISCUSSION

In this study, we analyzed multiple public data sets and applied machine learning techniques to identify 4 key genes associated with PBC: PSMB7, TBK1, SLC29A1, and CD244. Through GSEA enrichment analysis and immune cell infiltration assessment, we discovered that these genes may play significant roles in regulating immune responses, inflammation, antigen presentation, and metabolic activities, highlighting their diagnostic and therapeutic potential in PBC (11). We observed a marked increase in PSMB7 expression among patients with PBC, potentially affecting T-cell activation and immune tolerance by modulating protein degradation pathways. Its trypsin-like enzymatic activity could be crucial for antigen processing, thereby influencing immune responses (12). TANK functions as a critical modulator of the nuclear factor kappa-light-chain-enhancer of activated B cells signaling pathway (13), affecting inflammatory and antiviral defenses through the regulation of proteins such as TNF recepter-associated factor 2, TBK1, and inhibitor of nuclear factor kappa-B kinase subunit Epsilon (14). In PBC, elevated TANK levels may alter innate immune activation by modulating the nucleotide-binding oligomerization domain-like and retinoic acid-inducible gene-like receptor pathways, contributing to liver inflammation (15,16). SLC29A1, a nucleoside transporter, facilitates the transmembrane movement of purine and pyrimidine nucleosides, playing a pivotal role in cellular energy metabolism and controlling inflammatory pathways (17). Increased expression of SLC29A1 may exacerbate bile duct damage by disrupting immune cell metabolism and cytokine-mediated inflammatory signaling. CD244, which is expressed on natural killer (NK) and T cells and is involved in non-MHC-restricted cytotoxicity and T-cell activation (18,19), is expressed at elevated levels in the liver tissues of patients with PBC, particularly in portal areas with inflammatory infiltration.

Using the DGIdb database, we identified notable interactions between PSMB7, SLC29A1, and several approved anticancer agents, with CD244 showing a specific association with the drug ALDESLEUKIN. These results indicate that currently available immunomodulatory therapies could be beneficial for managing PBC (20). PSMB7, as an integral part of the proteasome complex, may modulate immune responses in PBC through its interaction with proteasome inhibitors such as bortezomib, potentially affecting antigen presentation mechanisms. The SLC29A1 gene plays a role in drug transport and cellular metabolism, which may affect bile acid pathways and influence the effectiveness of nucleoside analog-based therapies (21). In addition, CD244, a key receptor found on NK cells, interacts with ALDESLEUKIN, a drug known for its potential to enhance NK cell-mediated anti-inflammatory responses. These findings suggest that CD244 could serve as an immunotherapeutic target for PBC (18). These observations indicate that ALDESLEUKIN may have potential applications in PBC treatment. Although

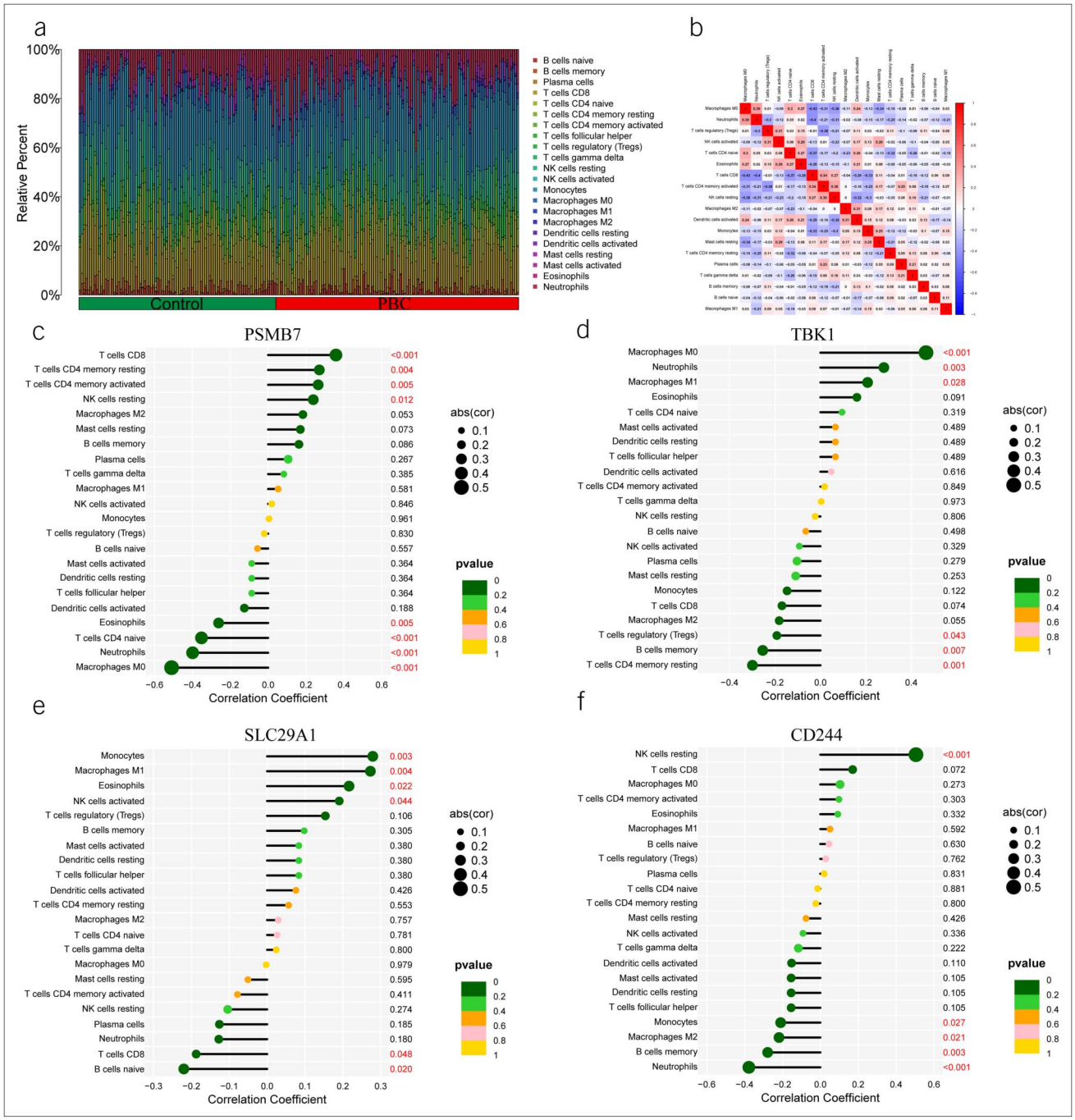

Figure 2. (**a**) Proportional composition of different immune-infiltrating cells in the control and primary biliary cholangitis (PBC) groups. (**b**) Correlation heatmap of immune-infiltrating cells in patients with PBC. Correlations between immune infiltration and the expression of key genes: (**c**) proteasome subunit beta type-7 (PSMB7), (**d**) TANK-binding kinase 1 (TBK1), (**e**) solute carrier family 29 member 1 (SLC29A1), and (**f**) natural killer cell receptor 2B4 (CD244).

this drug and similar agents have yet to receive approval for PBC management, their interactions with PBC-related genes open the door for further studies on immunomodulatory therapies and precision medicine. Although anticancer drugs are not part of the current standard treatment for PBC, their possible immunoregulatory properties warrant additional investigation. Future research may help determine whether these agents could contribute to the development of targeted therapeutic approaches for PBC.

This study established a ceRNA regulatory network, demonstrating that TBK1, CD244, and SLC29A1 modulate immune cell function through interactions with miRNAs and lncRNAs, playing pivotal roles in the immunopathogenesis of PBC. TBK1 is regulated by miR-365a-3p in our predicted ceRNA network. miR-365a-3p has been previously reported to regulate cell proliferation and migration in Hep-2 carcinoma cells by targeting ten-eleven translocation 1 (22), suggesting potential roles in cellular signaling. Although its specific function in immune cells or in the context of PBC has not been experimentally confirmed, it may potentially influence the PBC immune microenvironment through post-transcriptional regulation of TBK1.

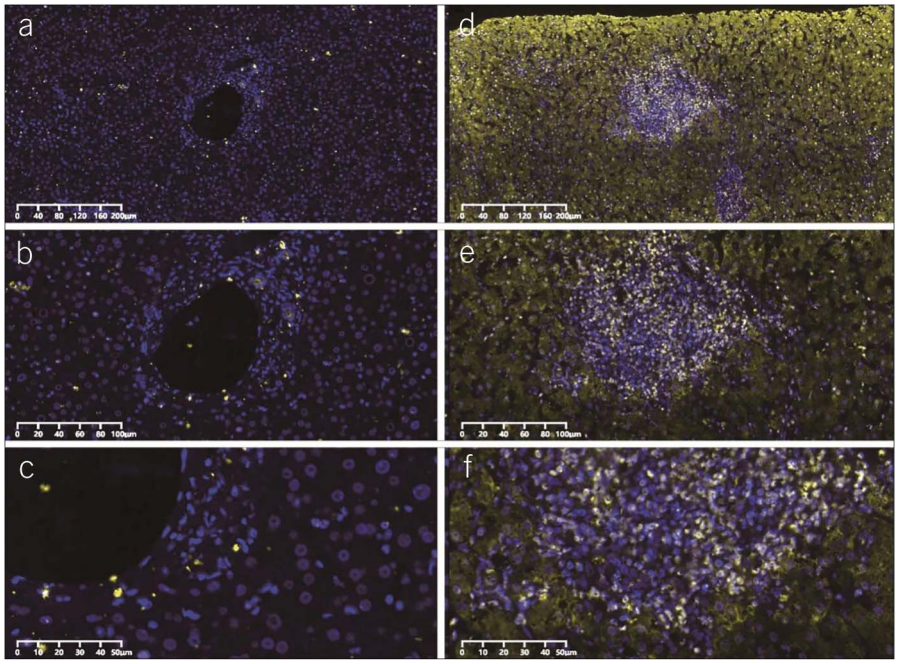

Figure 3. Immunofluorescence staining of proteasome subunit beta type-7 (PSMB7) in normal liver tissue (**a–c**) and primary biliary cholangitis (PBC) liver tissue (**d–f**). (**a, b**) Magnification ×10, (**c, d**) magnification ×20, (**e, f**) magnification ×40.

SLC29A1 is modulated by miR-185-5p and miR-149-3p, potentially affecting the energy supply and metabolism in immune cells (23). miR-149-3p participates in both the initiation and resolution of immune responses (24), and its interaction with SLC29A1 suggests that SLC29A1 may play a critical regulatory role in immune injury in PBC. CD244 is influenced by miR-876-5p and miR-26b-3p and further regulates NK cell function and immune surveillance through the lncRNA network. As an essential receptor on NK cells, CD244 may be integral to immune tolerance and evasion in PBC (25). This complex regulatory network offers novel insights into PBC immune regulation at the noncoding RNA level and provides a foundation for RNA-targeted therapies.

The novelty of this study lies in the combination of machine learning and experimental validation to systematically screen key genes for PBC and uncover their roles in the immune

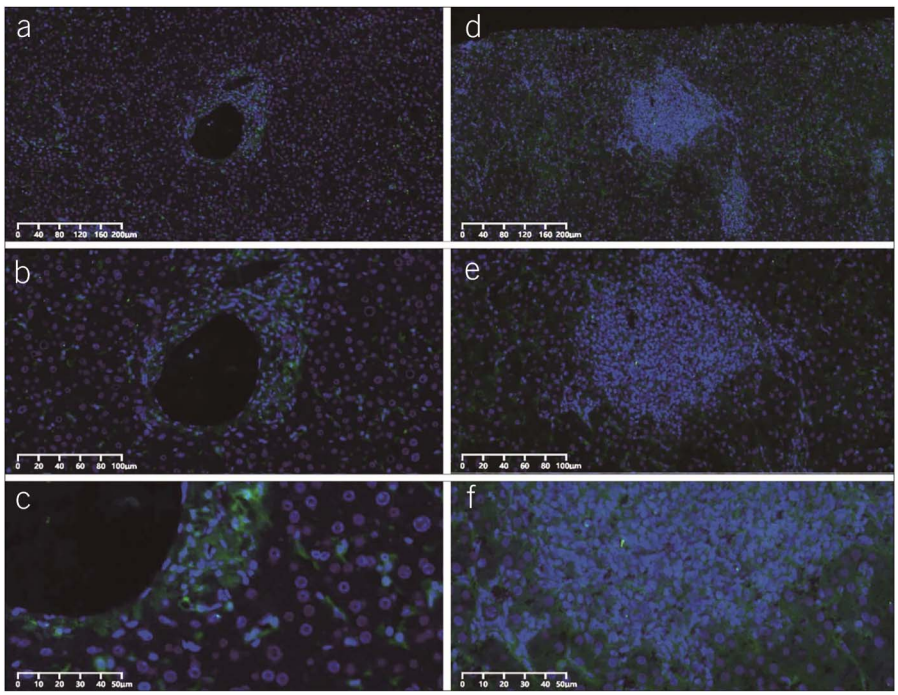

Figure 4. Immunofluorescence staining of TANK in normal liver tissue (**a–c**) and primary biliary cholangitis (PBC) liver tissue (**d–f**). (**a, b**) Magnification ×10, (**c, d**) magnification ×20, (**e, f**) magnification ×40.

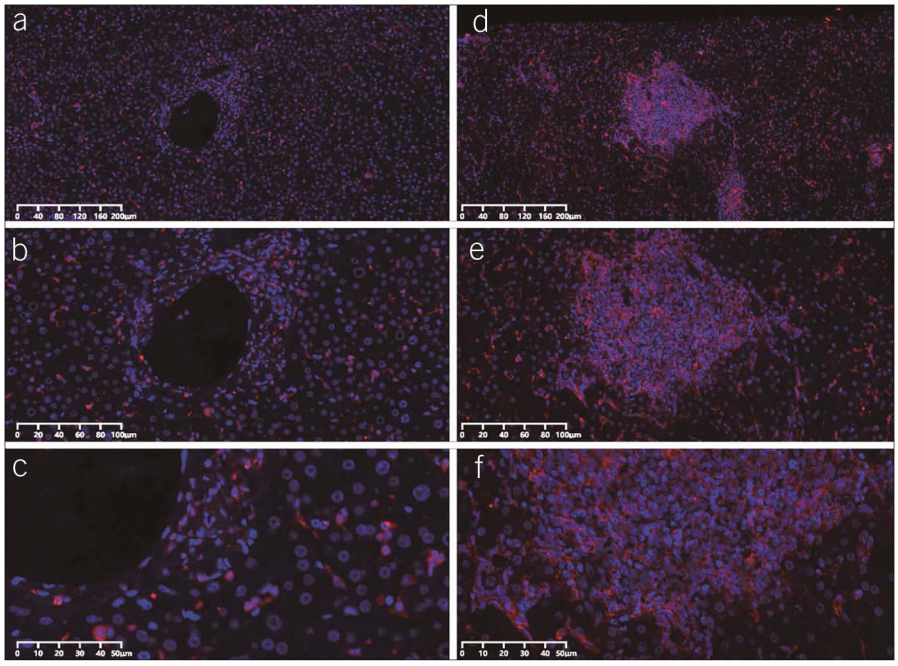

Figure 5. Immunofluorescence staining of equilibrative nucleoside transporter 1 (ENT1) in normal liver tissue (**a–c**) and primary biliary cholangitis (PBC) liver tissue (**d–f**). (**a, b**) Magnification ×10, (**c, d**) magnification ×20, (**e, f**) magnification ×40.

microenvironment. Compared with traditional methods based on individual gene analysis, this study integrated multiomics data and applied machine learning for more accurate and specific gene selection, offering new directions for early PBC diagnosis and targeted therapy. Nevertheless, this study has certain limitations. The Gene Expression Omnibus data sets used lack detailed immunoserological and clinical information, which are known to influence PBC progression and treatment response (26,27). Although we validated gene expression in liver tissues from 10 patients with PBC, the small sample size and lack of serum data precluded further stratified analysis. In addition, antinuclear antibody detection is serum-based and cannot be reliably assessed in tissue sections, limiting our ability to explore its association with gene expression. Future studies should address these limitations by integrating transcriptomic data with clinical and serological features in larger patient cohorts, and by combining

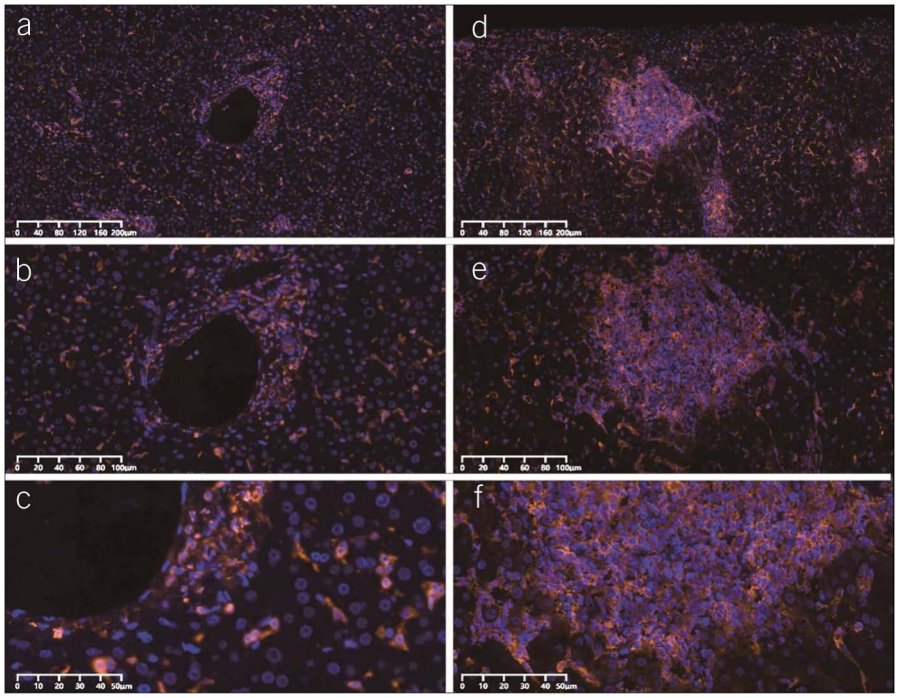

Figure 6. Immunofluorescence staining of natural killer cell receptor 2B4 (CD244) in normal liver tissue (**a–c**) and primary biliary cholangitis (PBC) liver tissue (**d–f**). (**a, b**) Magnification ×10, (**c, d**) magnification ×20, (**e, f**) magnification ×40.

single-cell sequencing and functional experiments to further elucidate the mechanisms, diagnostic relevance, and therapeutic potential of the identified genes in distinct PBC subtypes.

In this study, we identified 4 key genes—PSMB7, TBK1, SLC29A1, and CD244—using machine learning methods and explored their mechanisms in PBC through multiple analytical approaches. These results indicate that these genes may influence immune inflammation, metabolic regulation, and antigen presentation and are significantly upregulated in the liver tissues of patients with PBC, suggesting their potential diagnostic value. In addition, drug target prediction suggests that existing immunomodulators or anticancer drugs may be effective for PBC treatment, offering new directions for precision therapy. This study provides new insights into the molecular mechanisms, early diagnosis, and targeted treatment of PBC, although further studies are needed to validate their clinical applicability.

CONFLICTS OF INTEREST

Guarantor of the article: Longke Xie.

Specific author contributions: Z.Z.: writing—original draft, visualization, project administration, methodology, investigation, funding acquisition, formal analysis, data curation, and conceptualization. J.F.: writing—original draft, methodology, formal analysis, and data curation. L.C.: methodology, data curation. Y.L.: software, formal analysis. D.L.: writing–review and editing, visualization, project administration, conceptualization. L.X., L.X.: writing–review and editing, supervision, resources, project administration, data curation.

Financial support: This study was supported by grants from the Science and Technology Cooperation Program of Fujian Province (No. 2021I0036). Key discipline support projects of 900th Hospital of PLA Joint Logistic Support Force (No. 2023XKXH03).

Potential competing interests: None to report.

Data availability: The datasets generated during the current study are available from the corresponding author upon reasonable request.

Consent for publication: All the authors have read and approved the submission of the manuscript.

Study Highlights

WHAT IS KNOWN

- ✓ Primary biliary cholangitis (PBC) is a chronic autoimmune liver disease that gradually progresses, making early diagnosis and treatment challenging.

WHAT IS NEW HERE

- ✓ Machine learning identified 4 key genes (PSMB7, TBK1, SLC29A1, and CD244) for PBC.
- ✓ Immunofluorescence confirmed gene overexpression in PBC liver tissues.
- ✓ Drug target prediction suggested potential immunomodulatory therapies for PBC.

REFERENCES

1. Trivella J, John BV, Levy C. Primary biliary cholangitis: Epidemiology, prognosis, and treatment. Hepatol Commun 2023;7(6):e0179.
2. Gulamhusein AF, Hirschfield GM. Primary biliary cholangitis: Pathogenesis and therapeutic opportunities. Nat Rev Gastroenterol Hepatol 2020;17(2):93–110.
3. Khanna A, Leighton J, Lee Wong L, et al. Symptoms of PBC–Pathophysiology and management. Best Pract Res Clin Gastroenterol 2018;34-35:41–7.
4. Gershwin ME, Ansari AA, Mackay IR, et al. Primary biliary cirrhosis: An orchestrated immune response against epithelial cells. Immunol Rev 2000;174:210–25.
5. Houri I, Hirschfield GM. Primary biliary cholangitis: Pathophysiology. Clin Liver Dis 2024;28(1):79–92.
6. Tanaka A, Ma X, Takahashi A, et al. Primary biliary cholangitis. Lancet 2024;404(10457):1053–66.
7. van Hooff MC, Werner E, van der Meer AJ. Treatment in primary biliary cholangitis: Beyond ursodeoxycholic acid. Eur J Intern Med 2024;124:14–21.
8. Ostrowski J, Goryca K, Lazowska I, et al. Common functional alterations identified in blood transcriptome of autoimmune cholestatic liver and inflammatory bowel diseases. Sci Rep 2019;9(1):7190.
9. Lei L, Bruneau A, El Mourabit H, et al. Portal fibroblasts with mesenchymal stem cell features form a reservoir of proliferative myofibroblasts in liver fibrosis. Hepatology 2022;76(5):1360–75.
10. Horvath S, Erhart W, Brosch M, et al. Obesity accelerates epigenetic aging of human liver. Proc Natl Acad Sci U S A 2014;111(43):15538–43.
11. Lee D. Activators of the 26S proteasome when protein degradation increases. Exp Mol Med 2025;57(1):41–9.
12. Boes B, Hengel H, Ruppert T, et al. Interferon gamma stimulation modulates the proteolytic activity and cleavage site preference of 20S mouse proteasomes. J Exp Med 1994;179(3):901–9.
13. Runde AP, Mack R, S JP, et al. The role of TBK1 in cancer pathogenesis and anticancer immunity. J Exp Clin Cancer Res 2022;41(1):135.
14. Ujevic A, Knizkova D, Synackova A, et al. TBK1-associated adapters TANK and AZI2 protect mice against TNF-Induced cell death and severe autoinflammatory diseases. Nat Commun 2024;15(1):10013.
15. Zheng J, Shi W, Yang Z, et al. RIG-I-like receptors: Molecular mechanism of activation and signaling. Adv Immunol 2023;158:1–74.
16. Li X, Hu X, You H, et al. Regulation of pattern recognition receptor signaling by palmitoylation. iScience 2025;28(2):111667.
17. Yao SY, Ng AM, Cass CE, et al. Nucleobase transport by human equilibrative nucleoside transporter 1 (hENT1). J Biol Chem 2011;286(37):32552–62.
18. Sun L, Gang X, Li Z, et al. Advances in understanding the roles of CD244 (SLAMF4) in immune regulation and associated diseases. Front Immunol 2021;12:648182.
19. Nath PR, Maclean M, Nagarajan V, et al. Single-cell profiling identifies a CD8(bright) CD244(bright) Natural Killer cell subset that reflects disease activity in HLA-A29-positive birdshot chorioretinopathy. Nat Commun 2024;15(1):6443.
20. Rouette A, Trofimov A, Haberl D, et al. Expression of immunoproteasome genes is regulated by cell-intrinsic and -extrinsic factors in human cancers. Sci Rep 2016;6:34019.
21. Gao PT, Cheng JW, Gong ZJ, et al. Low SLC29A1 expression is associated with poor prognosis in patients with hepatocellular carcinoma. Am J Cancer Res 2017;7(12):2465–77.
22. Li J, Shen N, Bai GP, et al. MiR-365a-3p suppresses proliferation and invasion of Hep-2 cells through targeting ten-eleven translocation 1 (TET1). Neoplasma 2018;65(5):730–5.
23. Niemann B, Haufs-Brusberg S, Puetz L, et al. Apoptotic brown adipocytes enhance energy expenditure via extracellular inosine. Nature 2022;609(7926):361–8.
24. Zhang M, Gao D, Shi Y, et al. miR-149-3p reverses CD8(+) T-cell exhaustion by reducing inhibitory receptors and promoting cytokine secretion in breast cancer cells. Open Biol 2019;9(10):190061.
25. Mathew SO, Rao KK, Kim JR, et al. Functional role of human NK cell receptor 2B4 (CD244) isoforms. Eur J Immunol 2009;39(6):1632–41.
26. Granito A, Muratori P, Quarneti C, et al. Antinuclear antibodies as ancillary markers in primary biliary cirrhosis. Expert Rev Mol Diagn 2012;12(1):65–74.
27. Quarneti C, Muratori P, Lalanne C, et al. Fatigue and pruritus at onset identify a more aggressive subset of primary biliary cirrhosis. Liver Int 2015;35(2):636–41.