

Research Article

Identification of Six Diagnostic Biomarkers for Chronic Lymphocytic Leukemia Based on Machine Learning Algorithms

Yidong Zhu ¹, Xinjin Gan,² Ruoyan Qin,³ and Zhikang Lin³

¹Department of Traditional Chinese Medicine, Shanghai Tenth People's Hospital, Tongji University School of Medicine, Shanghai 200072, China

²Department of Hematology, Longhua Hospital, Shanghai University of Traditional Chinese Medicine, Shanghai 200032, China

³Department of Oncology, Longhua Hospital, Shanghai University of Traditional Chinese Medicine, Shanghai 200032, China

Correspondence should be addressed to Yidong Zhu; zhuyidong0813@foxmail.com

Received 17 July 2022; Revised 13 September 2022; Accepted 30 September 2022; Published 24 November 2022

Academic Editor: Shahid Mumtaz

Copyright © 2022 Yidong Zhu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Background. Chronic lymphocytic leukemia (CLL) is the most common type of leukemia in adults. Thus, novel reliable biomarkers need to be further explored to increase diagnostic, therapeutic, and prognostic effectiveness. **Methods.** Six datasets containing CLL and control samples were downloaded from the Gene Expression Omnibus database. Differential gene expression analysis, weighted gene coexpression network analysis (WGCNA), and the least absolute shrinkage and selection operator (LASSO) regression were applied to identify potential diagnostic biomarkers for CLL using R software. The diagnostic performance of the hub genes was then measured by the receiver operating characteristic (ROC) curve analysis. Functional analysis was implemented to uncover the underlying mechanisms. Additionally, correlation analysis was performed to assess the relationship between the hub genes and immunity characteristics. **Results.** A total number of 47 differentially expressed genes (DEGs) and 25 candidate hub genes were extracted through differential gene expression analysis and WGCNA, respectively. Based on the 14 overlapped genes between the DEGs and the candidate hub genes, LASSO regression analysis was used, which identified a final number of six hub genes as potential biomarkers for CLL: *ABCA6*, *CCDC88A*, *PMEPA1*, *EBF1*, *FILIP1L*, and *TEAD2*. The ROC curves of the six genes showed reliable predictive ability in the training and validation cohorts, with all area under the curve (AUC) values over 0.80. Functional analysis revealed an abnormal immune status in the CLL patients. A significant correlation was found between the hub genes and the immune-related pathways, indicating a possible tight connection between the hub genes and tumor immunity in CLL. **Conclusion.** This study was based on machine learning algorithms, and we identified six genes that could be possible CLL markers, which may be involved in CLL pathogenesis and progression through immune-related signal pathways.

1. Introduction

Chronic lymphocytic leukemia (CLL) is a hematopoietic malignancy characterized by the clonal accumulation of mature B lymphocytes in the peripheral blood, bone marrow, and lymphoid tissues [1]. It was estimated that 20,160 new cases would be diagnosed, and 4,410 cases of death of this disease would occur only in the United States in 2022 [2]. Notably, most newly diagnosed patients present only with asymptomatic peripheral blood lymphocytosis or leukocytosis [3]. Currently, the diagnosis of CLL is based mainly on blood counts, blood smears, and immunopheno-

typing of circulating B lymphocytes [4, 5]. Immunohistologically, the coexpression of CD5 and CD23 on the clonal population of B cells, detected by flow cytometry, can be used for the diagnosis of most CLL cases [6, 7]. Unfortunately, a risk of misdiagnosis might exist between CLL and other lymphoid malignancies with similar morphological features and CD5 positivity, such as mantle cell lymphoma [8, 9] and atypical lymphoplasmacytic lymphoma [10]. Neither the prognosis nor the treatment regimen of the aforementioned disorders is the same, leading to serious adverse consequences of a misdiagnosis [11–13]. Due to the overlapping immunophenotypes and the possibility of

the occurrence of similar histological patterns, lymphomas may resemble CLL. Hence, flow cytometry, immunohistochemistry, and anatomical pathology have particular limitations in the diagnosis of CLL from other entities [14, 15]. Moreover, CLL cases are not uniform in clinical practice, and some exhibit atypical features such as CD5- or CD23-negative [7, 16, 17]. Therefore, novel, more reliable biomarkers are needed to improve the diagnosis of CLL.

Over the past decade, next-generation sequencing technology and microarray analysis have been widely applied as fundamental methods in neoplastic disorders for multiple clinical utilization, including molecular diagnosis and prediction of prognosis [18, 19]. Genome studies have been conducted to explore the transcriptome changes in CLL, offering a novel method for potential markers and therapeutic targets [20, 21]. For instance, the *PTX3* gene was considered a marker associated with CLL disease [22]. Moreover, the oncogene *MSI2* was identified as a differential prediagnostic marker and potential therapeutic target of CLL [23]. In a previous study, miR-15b and miR-195 were reported to have the potential to function as novel and noninvasive biomarkers in the diagnosis and prognosis of patients with B-CLL [24]. However, these studies still have disadvantages as they are mainly single-center with small sample sizes, and their results have not been verified by large-scale clinical application.

It was demonstrated that the metabolic, protein interaction, and gene expression networks in biological environment fit in a scale-free topological distribution [25]. Genes are clustered in the form of a coexpression network, in which the ones connected with more genes are in the core position in modules with high modular identity, which are called hub genes [26, 27]. In previous studies, they have been distinguished by the gene expression difference of samples subjected to differential expression analysis alone [28, 29]. Weighted gene coexpression network analysis (WGCNA) is a systematic biology method which is focused on establishing the correlation patterns among genes across microarray samples and screening out hub genes without subjective judgement [30, 31]. The least absolute shrinkage and selection operator (LASSO) has a strong predictive value and low correlation for the selection of the best features for high-dimensional data and prevent overfitting during modeling [32, 33]. By its combination with the aforementioned bioinformatics analysis, in the present study, we aimed to identify genes as diagnostic biomarkers for CLL, with their differences and correlations. Additionally, we performed functional enrichment analysis to explore the possible mechanisms of their action and interaction.

2. Materials and Methods

2.1. Data Collection. Six datasets containing CLL and control samples (Table 1), GSE14853, GSE26725, GSE31048, GSE50006, GSE51528, and GSE55288, were downloaded from the Gene Expression Omnibus database (GEO, <https://www.ncbi.nlm.nih.gov/geo/>). The raw data of all studied datasets were normalized to eliminate batch effects. We merged GSE14853, GSE26725, GSE50006, and GSE55288 as a training

TABLE 1: Characteristics of the studied datasets.

GEO series	Control samples	CLL samples	Data type
GSE14853	3	34	Training cohort
GSE26725	5	12	Training cohort
GSE31048	33	188	Validation cohort
GSE50006	32	188	Training cohort
GSE51528	12	217	Validation cohort
GSE55288	33	59	Training cohort

cohort for subsequent analysis, whereas GSE31048 and GSE51528 were selected as diverse validation cohorts to verify the result.

2.2. Identification of Differentially Expressed Genes (DEGs). CLL and control samples were subjected to differential expression analysis using the LIMMA package. DEGs had to conform to the criterion of $|\log FC| > 2$, where FC denotes fold change and adjusted $P < 0.05$. The heatmap and volcano plot of the DEGs were visualized by “pheatmap” and “ggplot2” packages.

2.3. Construction of a Gene Coexpression Network. WGCNA was performed on the training cohort to construct a gene coexpression network of CLL. Based on a scale-free topology model, we quantified and integrated goodness of fit with mean connectivity to ascertain the optimal soft threshold. Multiple modules were next detected automatically, and the topological overlap measure was computed to estimate the adjacencies and similarities among the different modules by average hierarchical clustering. Then, the topologically similar modules were combined into a new cluster. The Pearson correlation analysis was performed to assess the correlations between the module genes and the clinical characteristics. The module with the highest correlations was selected, and the genes in the obtained module were further assessed by Gene Significance (GS) and Module Membership (MM). The genes with an absolute value of MM of over 0.8 and GS of over 0.5 were considered as candidate hub genes for CLL. These measurements were analyzed and visualized by “LIMMA” and “WGCNA” packages.

2.4. Identification and Validation of the Hub Genes. The genes overlapped between the DEGs and the candidate hub genes from WGCNA were aggregated by the “venn” package. LASSO regression was adopted to identify the optimal panel of the final hub genes. Further, receiver operating characteristic (ROC) curves were constructed to investigate the accuracy and specificity of these genes for CLL diagnosis. Furthermore, the diagnostic value of the hub genes was verified in the validation cohorts (GSE31048 and GSE51528) by differential expression analysis and ROC curves. The results of these analyses were visualized by “ggpubr” and “pROC” packages.

2.5. Functional Enrichment and Correlation Analyses. To investigate the possible mechanisms, the DEGs were subjected to Gene Ontology (GO) and Kyoto Encyclopedia of

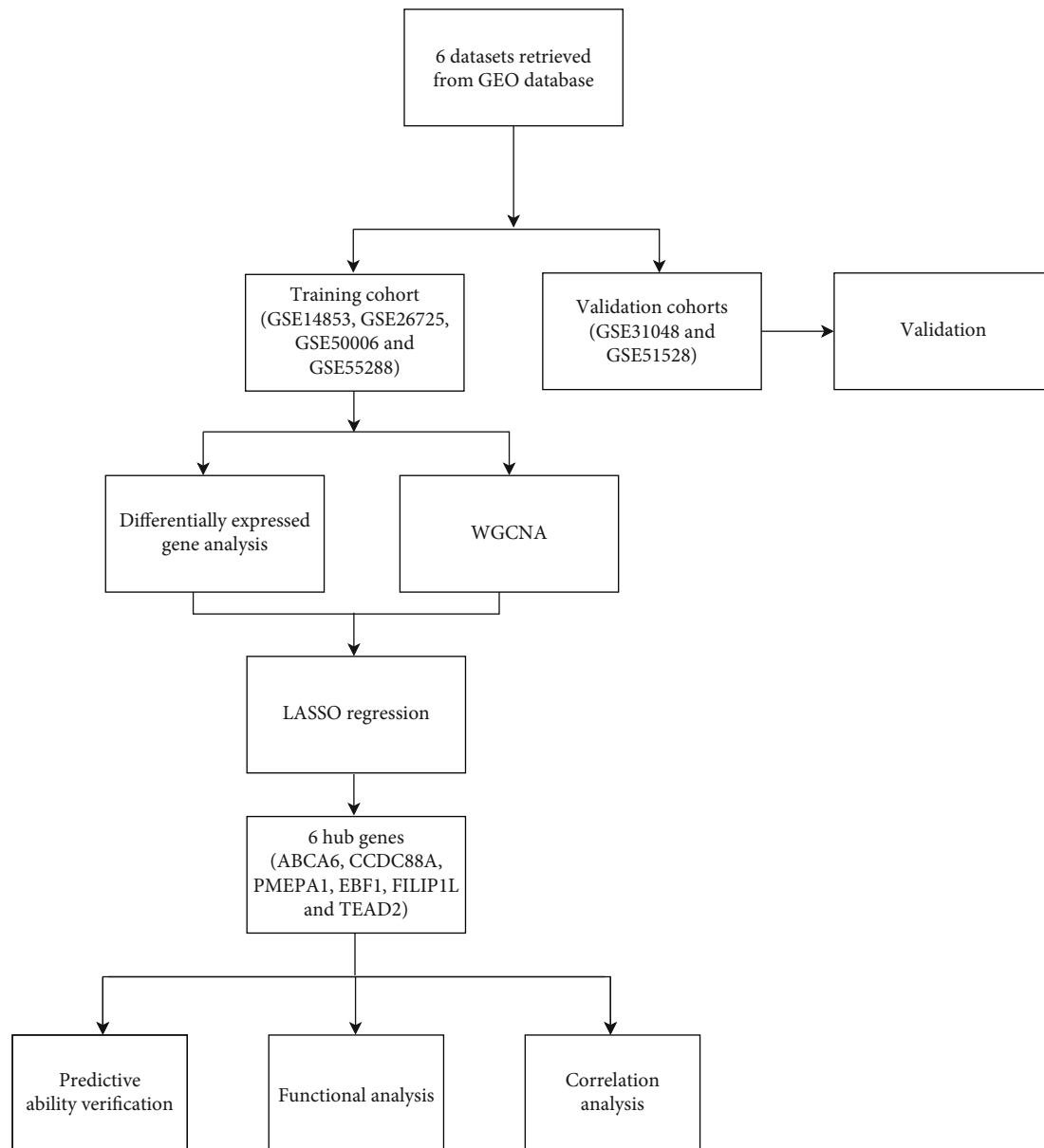


FIGURE 1: Flowchart of this study.

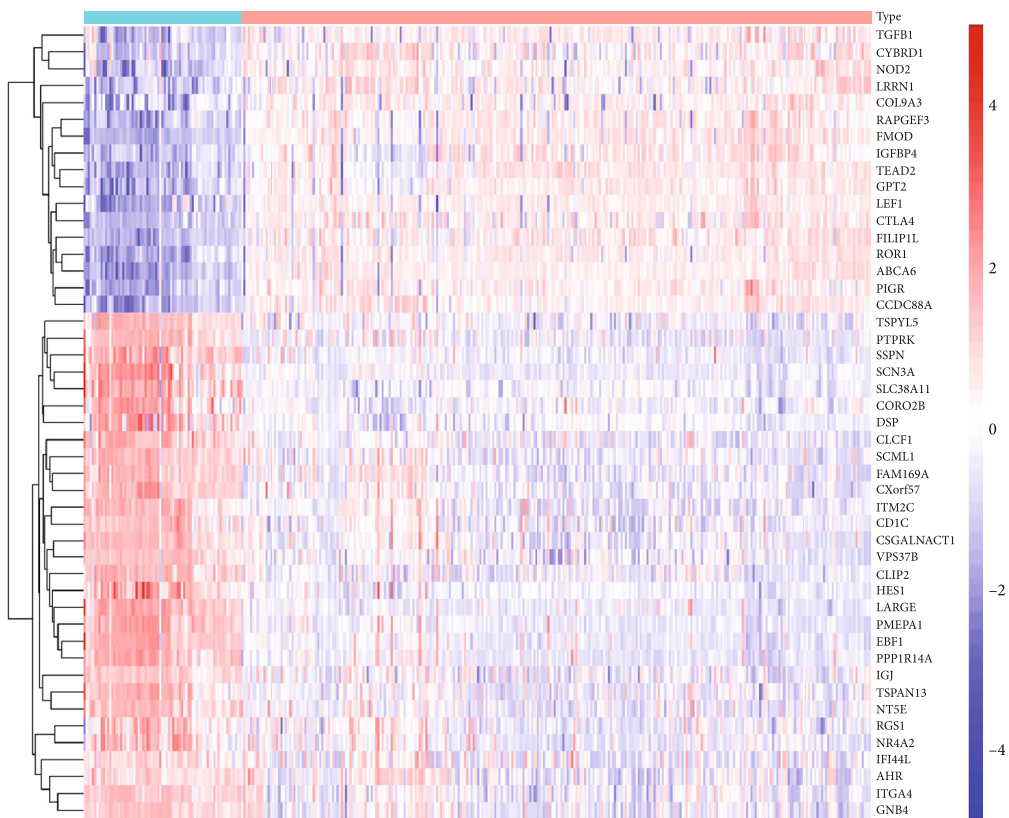
Genes and Genome (KEGG) analyses. In addition, single sample gene set enrichment analysis (ssGSEA) was applied to contrast the enrichment levels of the immune-related functions and cells between the CLL and the control samples. The Spearman correlation analysis was implemented to examine comprehensively the relationship between hub genes and immunity characteristics. The data were evaluated and visualized by the “clusterProfiler”, “enrichplot”, “DOSE”, “pheatmap”, “GSVA”, “GSEABase”, “vioplot”, and “tidyverse” packages.

2.6. Statistical Analysis. Data analysis was performed using R software (version 4.1.3). Student’s *t*-test was conducted to determine the significance of the discrepancy between the CLL and the control samples. Two-sided $P < 0.05$ was considered to indicate a statistically significant difference.

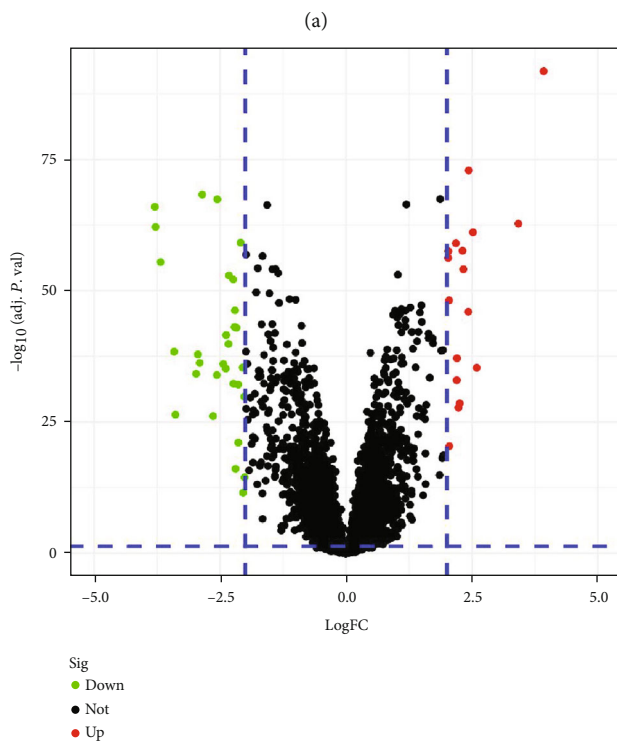
3. Results

3.1. Identification of DEGs. The flowchart illustrated in Figure 1 presents the identification and validation of the hub genes as well as the results of the subsequent analyses conducted in the present study. According to the criterion on the training set, 47 DEGs were identified between CLL and control samples, of which 17 were upregulated and 30 were downregulated in CLL (Figures 2(a) and 2(b)).

3.2. Construction of a Gene Coexpression Network. With the soft threshold of $\beta = 5$, the network was closer to the real biological network state as it adhered to the power law distribution (Figures 3(a) and 3(b)). A hierarchical clustering analysis based on weighted correlation was further applied. By segmenting the clustering results and merging the similar



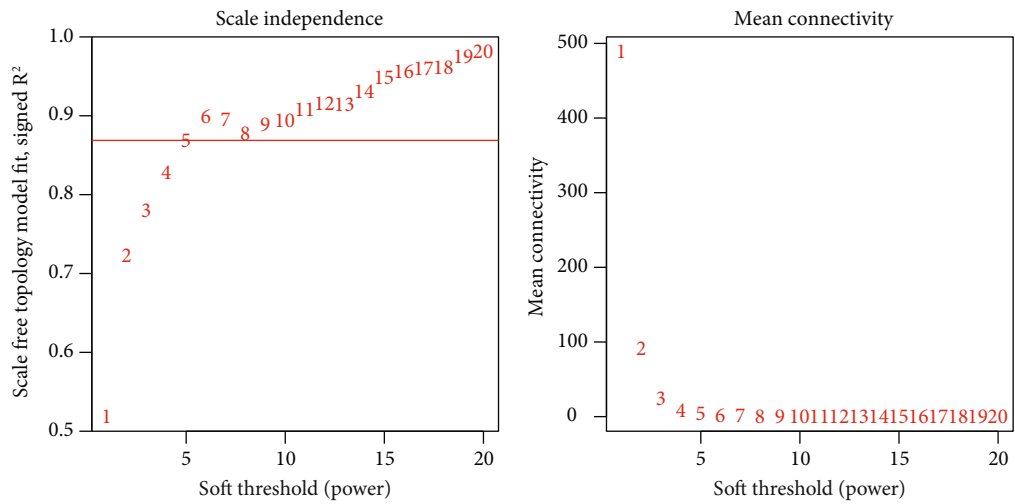
Type
 Con
 Treat



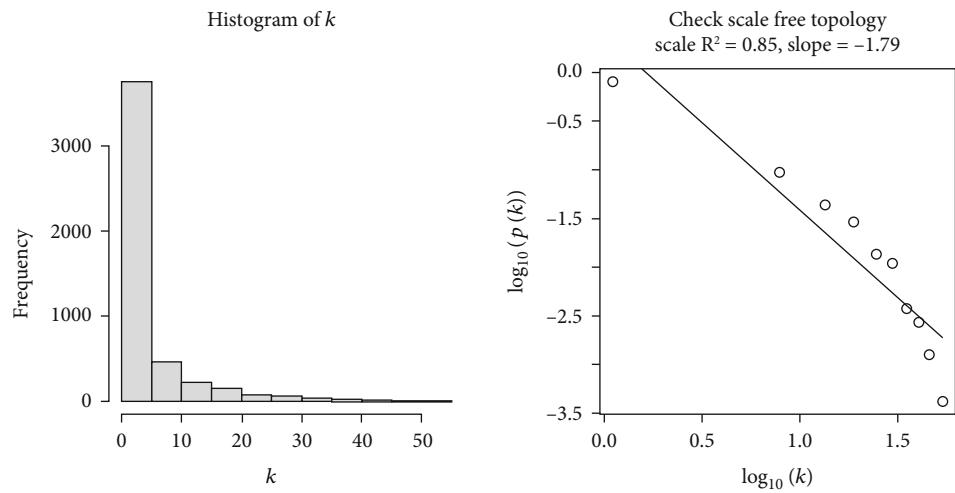
Sig
 ● Down
 ● Not
 ● Up

(b)

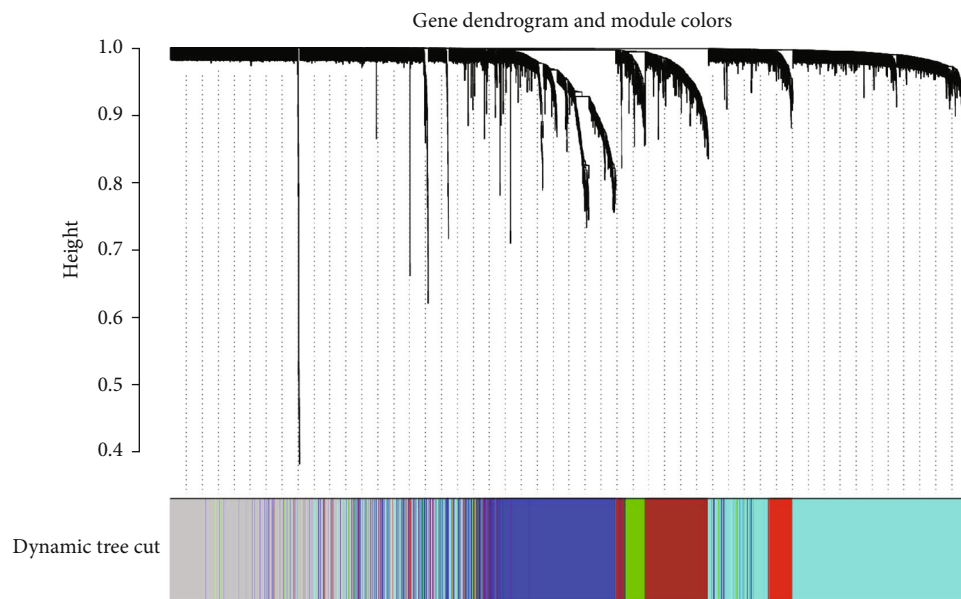
FIGURE 2: Identification of DEGs between the CLL and the control samples. (a) Heatmap of the DEGs; (b) volcano plot of the DEGs.



(a)

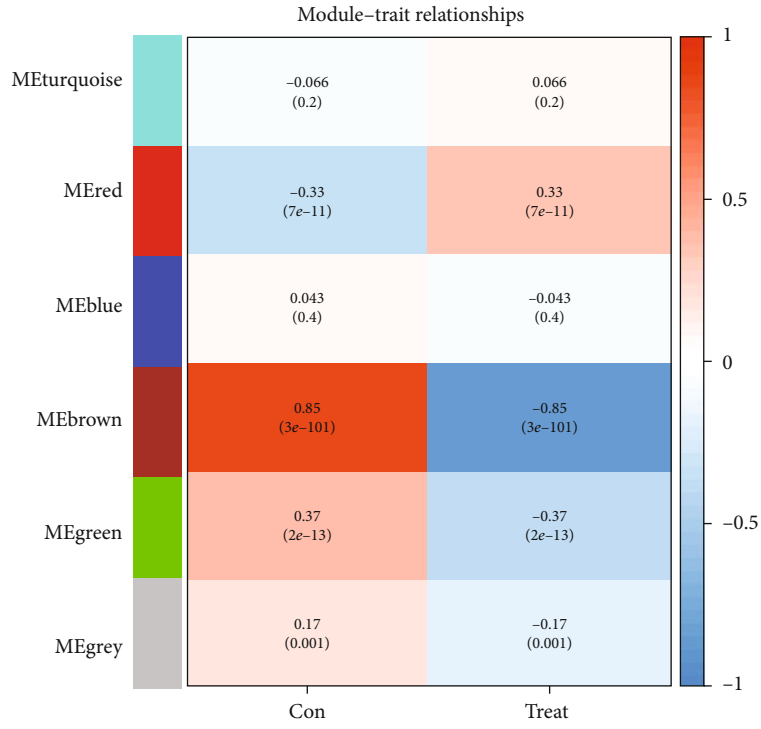


(b)



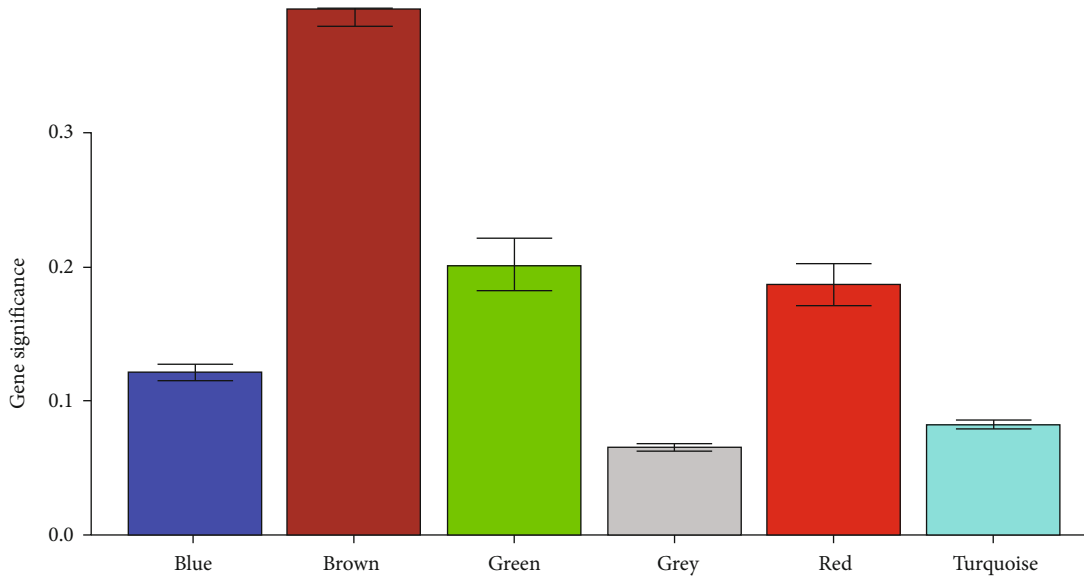
(c)

FIGURE 3: Continued.



(d)

Gene significance across modules, p -value = 0



(e)

FIGURE 3: Continued.

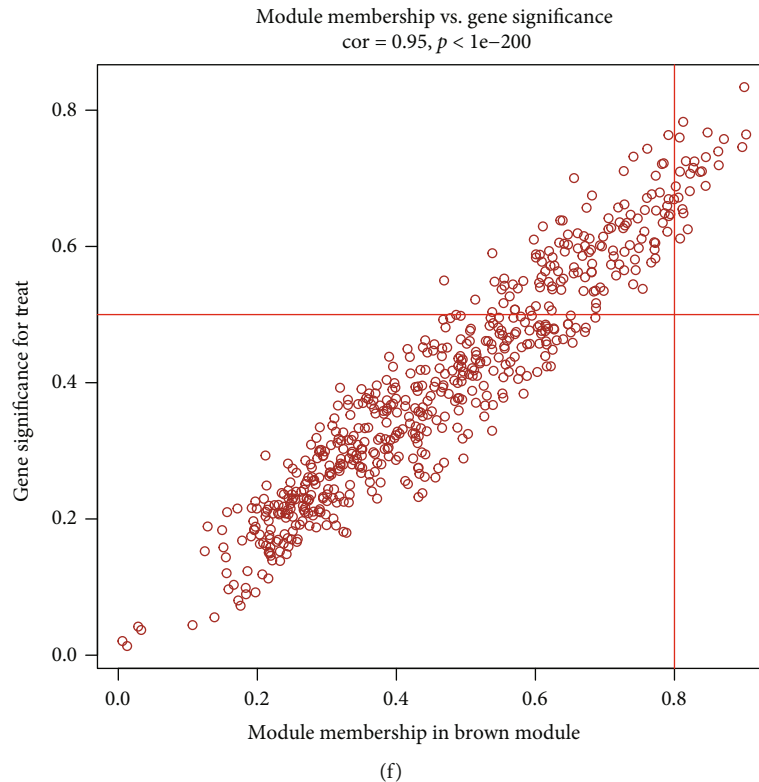


FIGURE 3: Construction of a gene coexpression network. (a) Analysis of the scale-free network topology for the optimal soft threshold; (b) validation of the optimal soft threshold; (c) WGCNA cluster dendrogram and module assignment; (d) heatmap of the module-trait relationships; (e) histogram of the gene significance across modules; (f) scatterplot of the module membership and gene significance in the brown module.

modules, a total number of six modules were obtained (Figures 3(c)). Of them, the brown module, including 589 genes, showed the strongest correlation with CLL ($cor = 0.85$, $P = 3e - 101$; Figures 3(d) and 3(e)). A tight correlation between GS and MM ($cor = 0.95$; $P < 1e - 200$) was also established. Based on the specified earlier criteria (an absolute value of $GS > 0.50$ and $MM > 0.80$), 25 genes in the brown module were identified as candidate hub genes, which were further subjected to subsequent analysis (Figure 3(f)).

3.3. Identification of the Hub Genes. Fourteen overlapped genes were obtained based on the intersections of the DEGs and the candidate hub genes (Figure 4(a)). To prevent overfitting and enhance the accuracy of the diagnostic value, LASSO analysis was utilized to extract the following six finally identified hub genes: *ABCA6*, *CCDC88A*, *PMEPA1*, *EBF1*, *FILIP1L*, and *TEAD2* (Figures 4(b) and 4(c)).

3.4. Characterization and Validation of the Hub Gene Expression and Diagnostic Value. The boxplots revealed the differential expression of the six hub genes between the CLL and control samples in the training cohort (Figures 5(a)–5(f)). Among them, *ABCA6*, *CCDC88A*, *FILIP1L*, and *TEAD2* were significantly upregulated, whereas *PMEPA1* and *EBF1* were significantly downregulated in the CLL samples. The results were verified in the validation cohorts, and the consistent gene expression

patterns were obtained (Figures 5(g)–5(r)). All area under the curve (AUC) values of the six hub genes were over 0.95 in the training cohort (Figures 6(a)–6(f)), showing a satisfactory diagnostic value for CLL. To verify the diagnostic value of the hub genes, we also constructed ROC curves for the validation sets (GSE31048 and GSE51528). The same result was obtained for the GSE31048 dataset, with all AUC values over 0.95 (Figures 6(g)–6(l)). In the GSE51528 dataset, the maximum AUC value was 0.815 for *FILIP1L* (Figure 6(m)). *CCDC88A* (Figure 6(n)) and *PMEPA1* (Figure 6(o)) had AUC values of 0.856 and 0.861, respectively, whereas the other three genes had AUC values over 0.95 (Figures 6(p)–6(r)).

3.5. Functional Enrichment and Correlation Analyses. Functional enrichment analysis was used to establish the possible mechanism involved in the genesis of CLL. GO function analysis results indicated that the biological processes included mainly B-cell proliferation, leukocyte cell-cell adhesion, and positive regulation of leukocyte cell-cell adhesion (Figure 7(a)). The KEGG analysis included predominantly an intestinal immune network for IgA production, ECM-receptor interaction, and hematopoietic cell lineage (Figure 7(b)). The ssGSEA results showed that various immune-related functions and cells had significant differences between the CLL and the control samples (Figures 7(c) and 7(d)). Furthermore, the correlation analysis revealed significant correlations between the hub genes

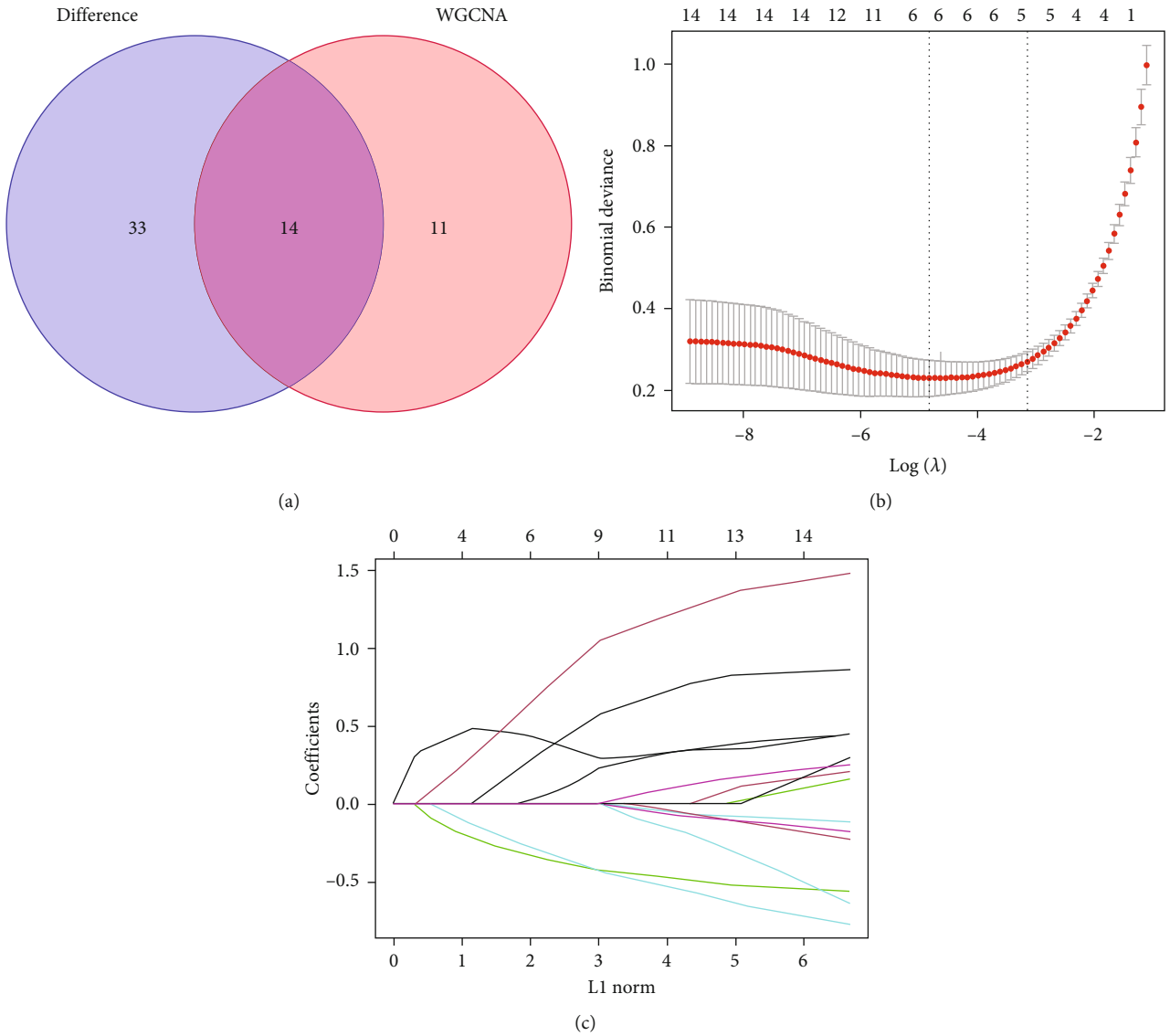


FIGURE 4: Identification of the hub genes. (a) Venn diagram of the DEGs and the candidate hub genes; (b) and (c) LASSO analysis of the hub genes.

and the immune-related functions and cells (Figures 7(e) and 7(f)), indicating that hub genes may play a crucial role in the pathogenesis and progression of CLL through immune status regulation.

4. Discussion

The primary finding of the present study is the successful identification and validation of six diagnostic biomarkers for CLL based on differentially expressed gene analysis, WGCNA, and LASSO regression. The following functional enrichment analysis revealed the crucial impact of immune dysfunction on CLL occurrence and progression. The significant correlations between hub genes and tumor immunity promote the uncovering of the possible mechanism of the involvement of these six genes in the pathogenesis of CLL. The identification of the diagnostic biomarkers might further facilitate the diagnosis of CLL in clinical practice.

In the present study, 47 DEGs and 25 candidate hub genes were extracted through differentially expressed gene analysis and WGCNA, respectively. Subsequently, LASSO regression analysis was used to identify the six final hub genes from the fourteen intersecting genes between the DEGs and the candidate hub genes. The gene expression patterns in the training cohort and different validation cohorts were consistent, with a statistically significant difference between the CLL and the control samples. Nevertheless, the values of control samples in the GSE51528 dataset were not the same as those in other cohorts. Unlike other datasets using normal cells isolated from peripheral blood as normal controls, the GSE51528 dataset used normal B-cell subpopulations from tonsils as controls. The expression of the same gene could be different in various tissues, and the AUC value could have differed because of the use of different tissues as controls. Despite this drawback, the AUC values of the six genes were over 0.8. The six newly identified biomarkers still

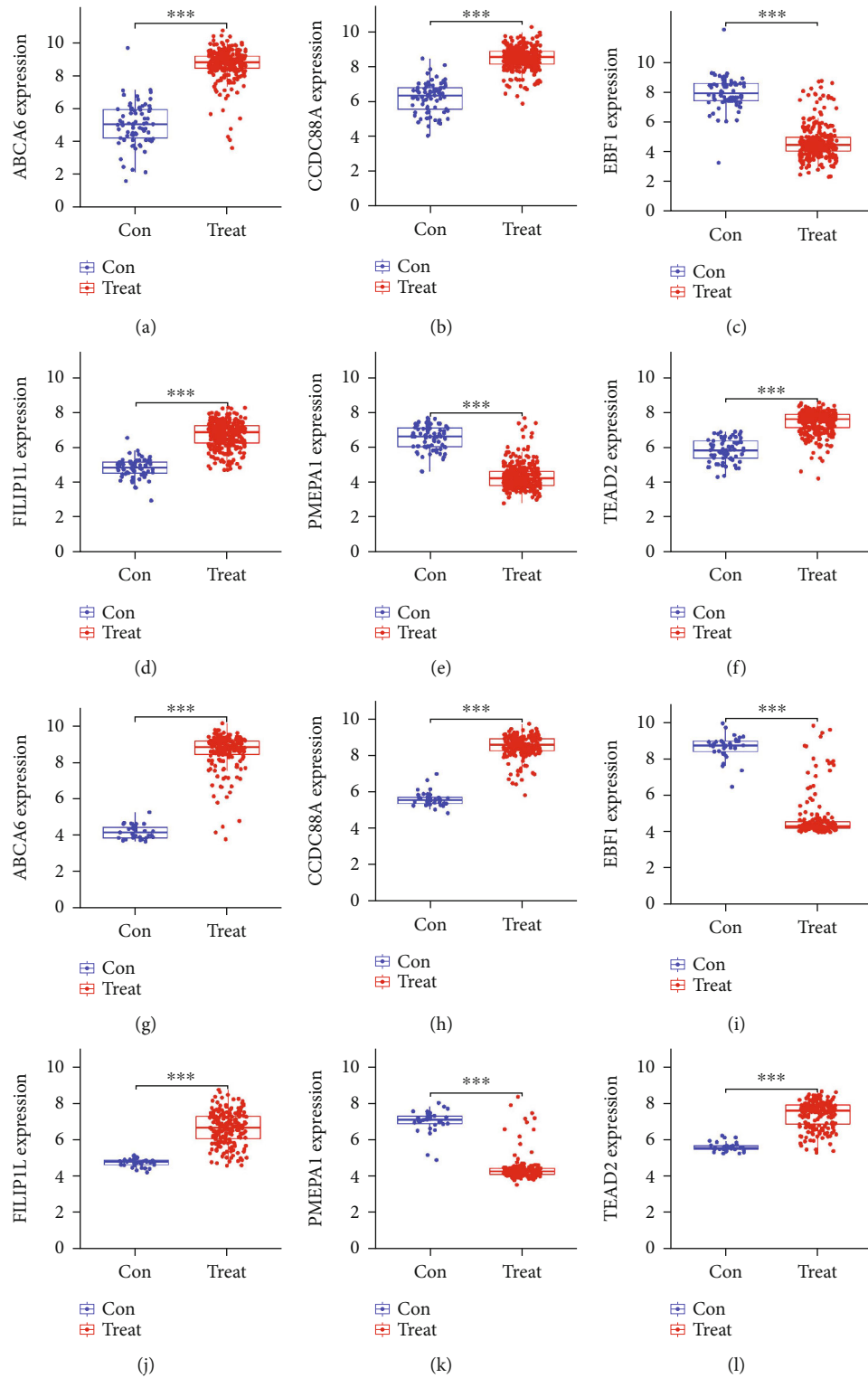


FIGURE 5: Continued.

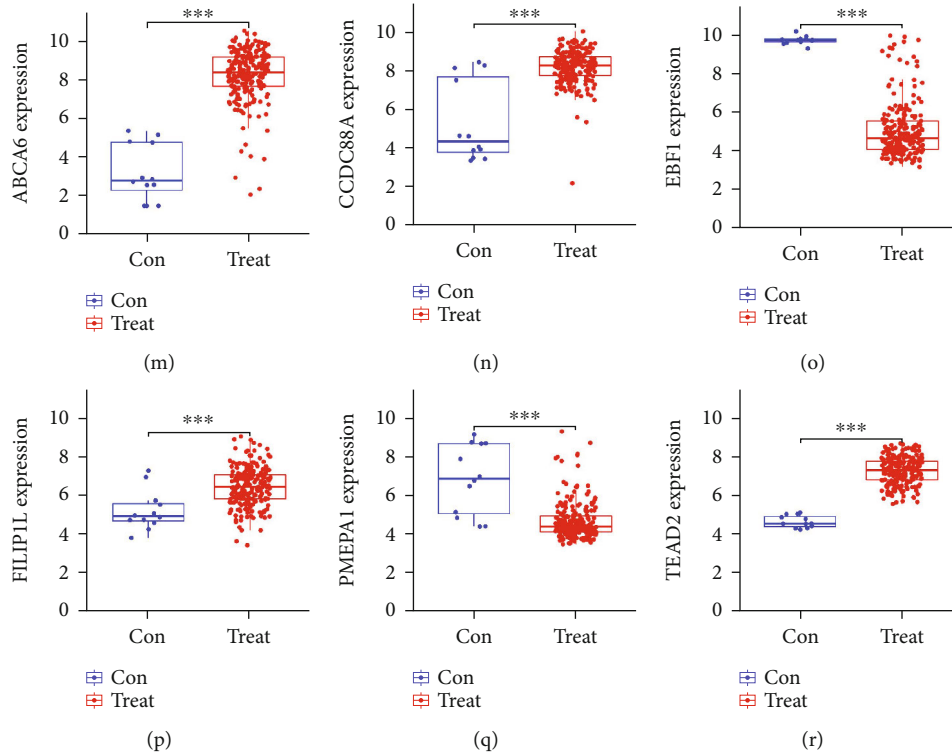


FIGURE 5: Differential hub gene expression between the CLL and the control samples. (a)–(f) *ABCA6* (a), *CCDC88A* (b), *EBF1* (c), *FILIP1L* (d), *PMEPA1* (e), and *TEAD2* (f) in the training cohort; (g)–(l) *ABCA6* (g), *CCDC88A* (h), *EBF1* (i), *FILIP1L* (j), *PMEPA1* (k), and *TEAD2* (l) in the GSE31048 dataset. (m)–(r) *ABCA6* (m), *CCDC88A* (n), *EBF1* (o), *FILIP1L* (p), *PMEPA1* (q), and *TEAD2* (r) in the GSE51528 dataset.

had a good performance in predicting CLL diagnosis even in the tonsils as controls instead of in specimens of peripheral blood. Of the identified hub genes, *ABCA6*, a member of the ATP-binding cassette transporter family, acts as a probable transporter which may play a role in macrophage lipid transport and homeostasis [34, 35]. *PMEPA1* regulates cell proliferation, differentiation, migration, and immunosuppression through the TGF-beta signaling pathway [36, 37]. *EBF1* has been confirmed to be the key pioneer transcription factor of B-cell specification and commitment by poising or activating lineage-specific genes and repressing genes related with alternative cell fates [38–40]. Overexpressed *FILIP1L* was found to modulate the antiangiogenic activity in endothelial cells, leading to inhibition of cell proliferation and migration as well as an increase in apoptosis [41]. *TEAD2* has been found to be relevant to tumor suppression by restricting proliferation and promoting apoptosis through the Hippo signaling pathway [42, 43]. *CCDC88A* serves as a nonreceptor guanine nucleotide exchange factor which binds to and activates guanine nucleotide-binding protein G (i) alpha subunits, involved in multiple biological processes, such as cell migration and cellular immunity [44–48]. However, the specific role of these six genes in CLL has not been reported yet.

To explore the possible mechanisms by which the hub genes were involved in the nosogenesis of CLL, we performed GO, KEGG, ssGSEA, and correlation analyses. GO function analysis indicated that the biological processes of

DEGs were enriched mainly in B-cell activation and proliferation as well as the regulation of leukocyte cell-cell adhesion. KEGG analysis predominantly included immune-associated pathways (e.g., an intestinal immune network for IgA production, ECM-receptor interaction, and leukocyte transendothelial migration). These results suggest that CLL is closely related to tumor immunity. Therefore, ssGSEA was applied to further investigate the differences in the immune cell infiltration between CLL and control samples. We found that T-cell coinhibition was significantly upregulated in the CLL group, which is a vital element contributing to immune function suppression by providing inhibitory signals to activated T cells [49, 50]. Furthermore, various immune-related cells, such as CD8+ T cells and regulatory T cells (Tregs), were significantly more overexpressed in the CLL samples than in the controls. CD8+ T cells exhibited profound functional deficits in the proliferation and cytotoxicity despite an increase in their absolute numbers in the peripheral blood [51, 52]. Treg, which is a subset of CD4+ T cells, was also found to be increased in CLL patients [53–55]. Dysregulation of Tregs leads to an unbalanced immune system and contributes to immune suppression, disease progression, and poor prognosis [56–58]. Coinhibitory molecules expressed by tumor infiltrating T cells and Tregs were previously considered to constitute a pivotal mechanism by promoting tumor immune evasion [59, 60]. The correlations of the hub genes identified here with immune-related functions and cells were also determined, showing significant

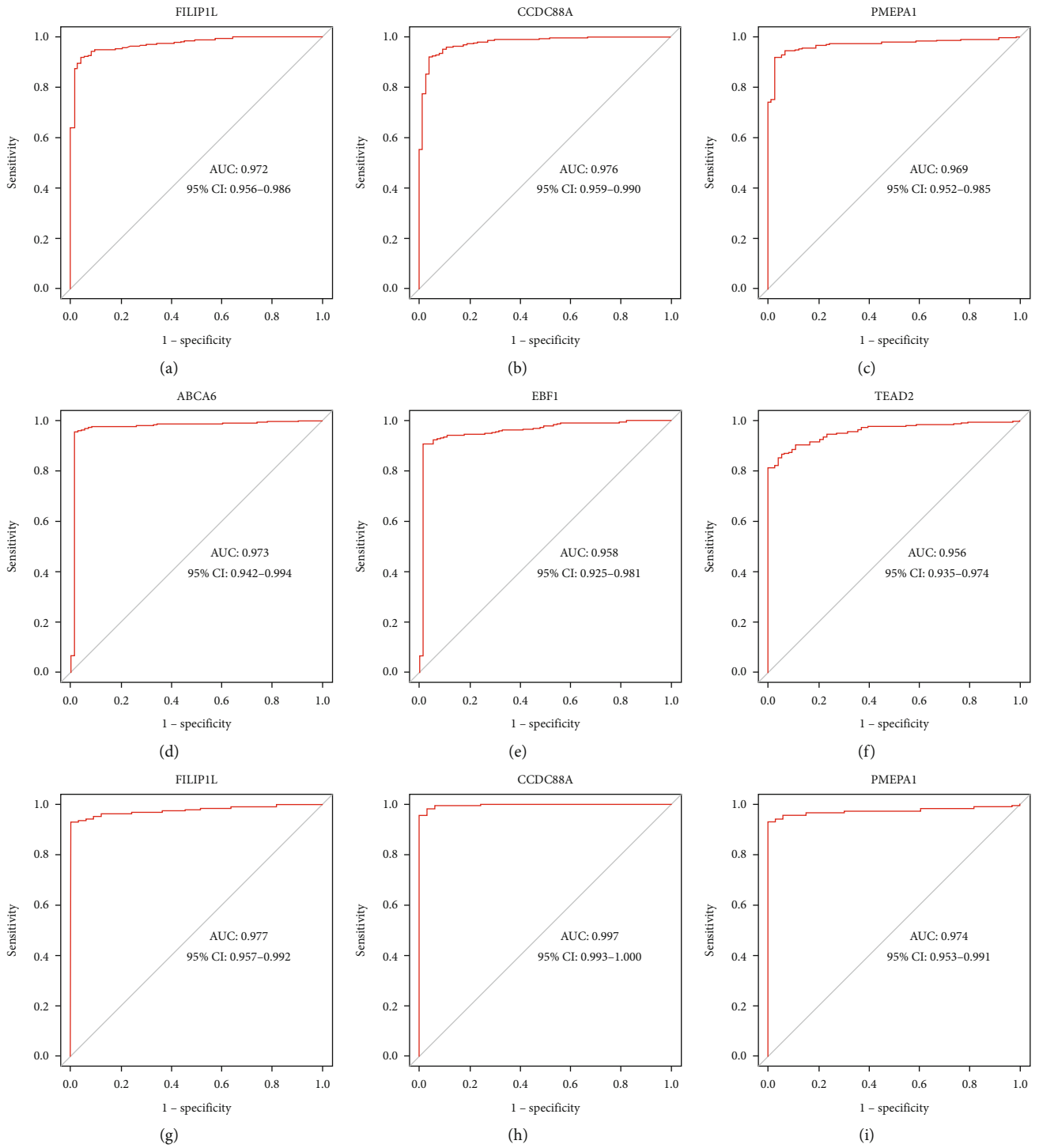


FIGURE 6: Continued.

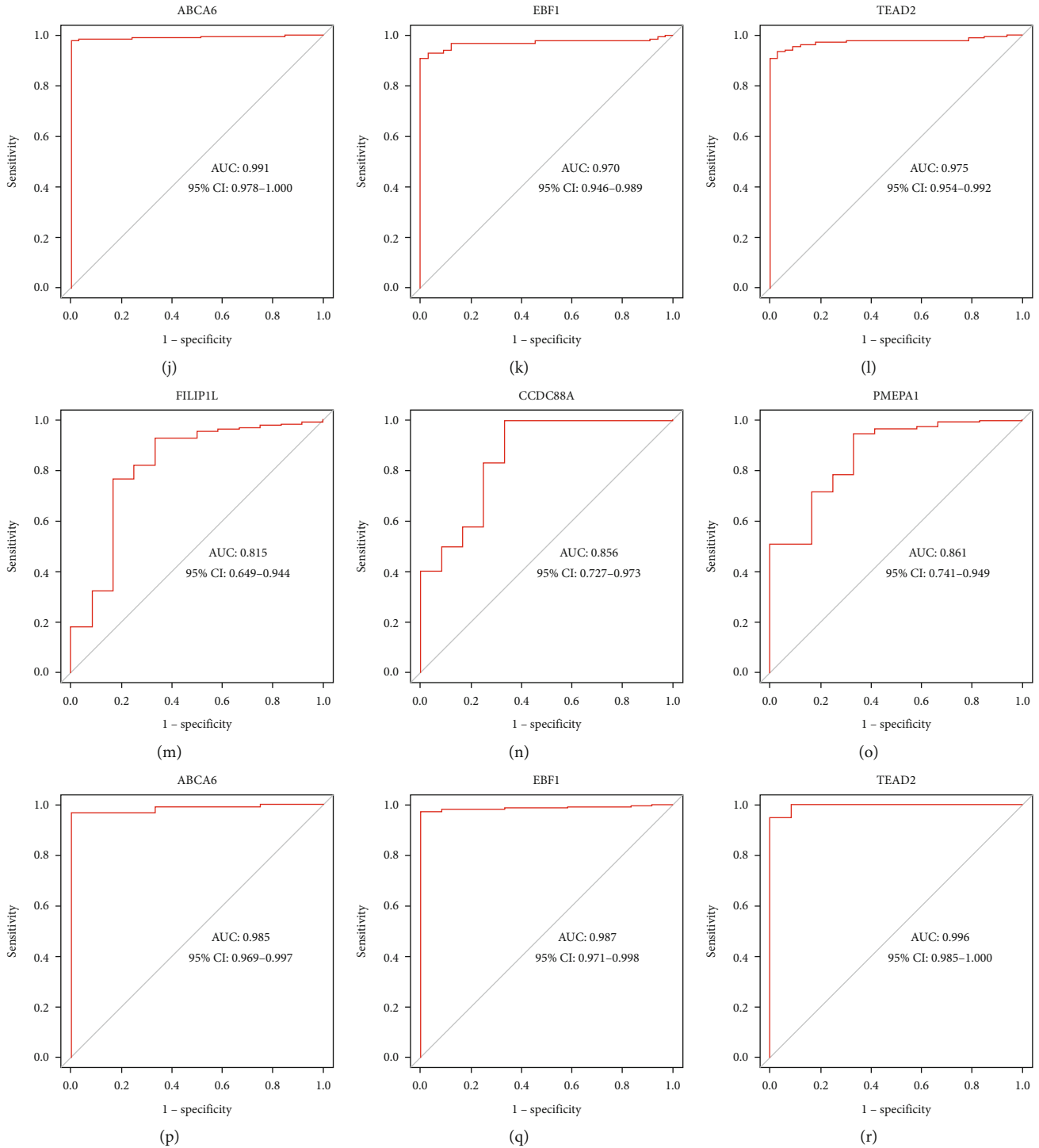
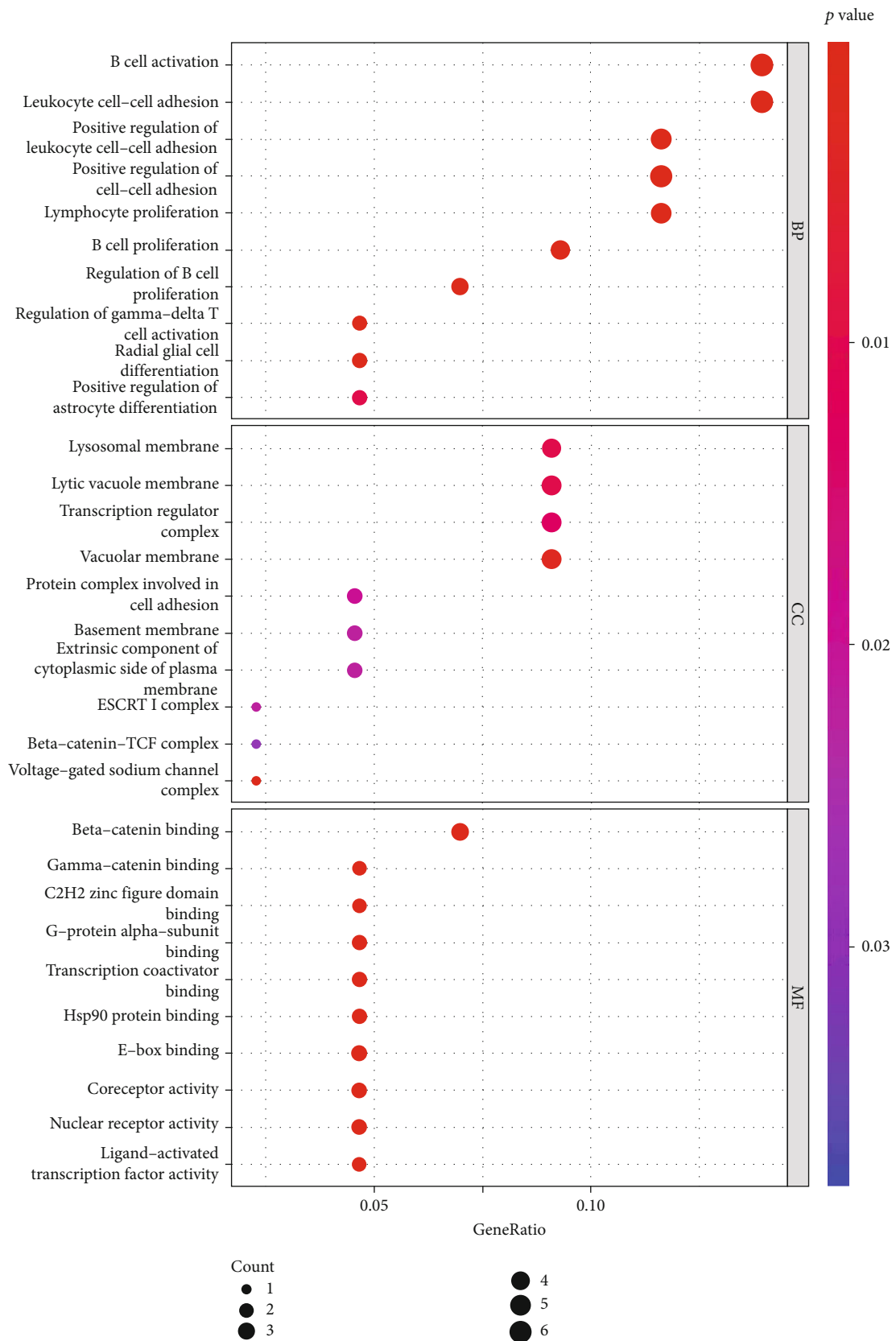


FIGURE 6: Diagnostic value of the hub genes for CLL. (a)–(f) ROC curves of *FILIP1L* (a), *CCDC88A* (b), *PMEPA1* (c), *ABCA6* (d), *EBF1* (e), and *TEAD2* (f) in the training cohort; (g)–(l) ROC curves of *FILIP1L* (g), *CCDC88A* (h), *PMEPA1* (i), *ABCA6* (j), *EBF1* (k), and *TEAD2* (l) in the GSE31048 dataset; (m)–(r) ROC curves of *FILIP1L* (m), *CCDC88A* (n), *PMEPA1* (o), *ABCA6* (p), *EBF1* (q), and *TEAD2* (r) in the GSE51528 dataset.

interactions between the hub genes and tumor immunity. Thus, it is reasonable to assume that the selected genes may modulate the tumor microenvironment and promote tumor immune evasion in CLL through inhibitory immune cells. Serious defects in the immune system and the capacity of leukemic cells to escape immune recognition played an

irreplaceable role in the nosogenesis and progression of CLL, which was in agreement with the findings of the present study [1, 61]. Therefore, we speculated that the biomarkers established may be involved in the suppression of the immune response, which supports the proliferation and survival of CLL cell through its immune-related pathways.



(a)

FIGURE 7: Continued.

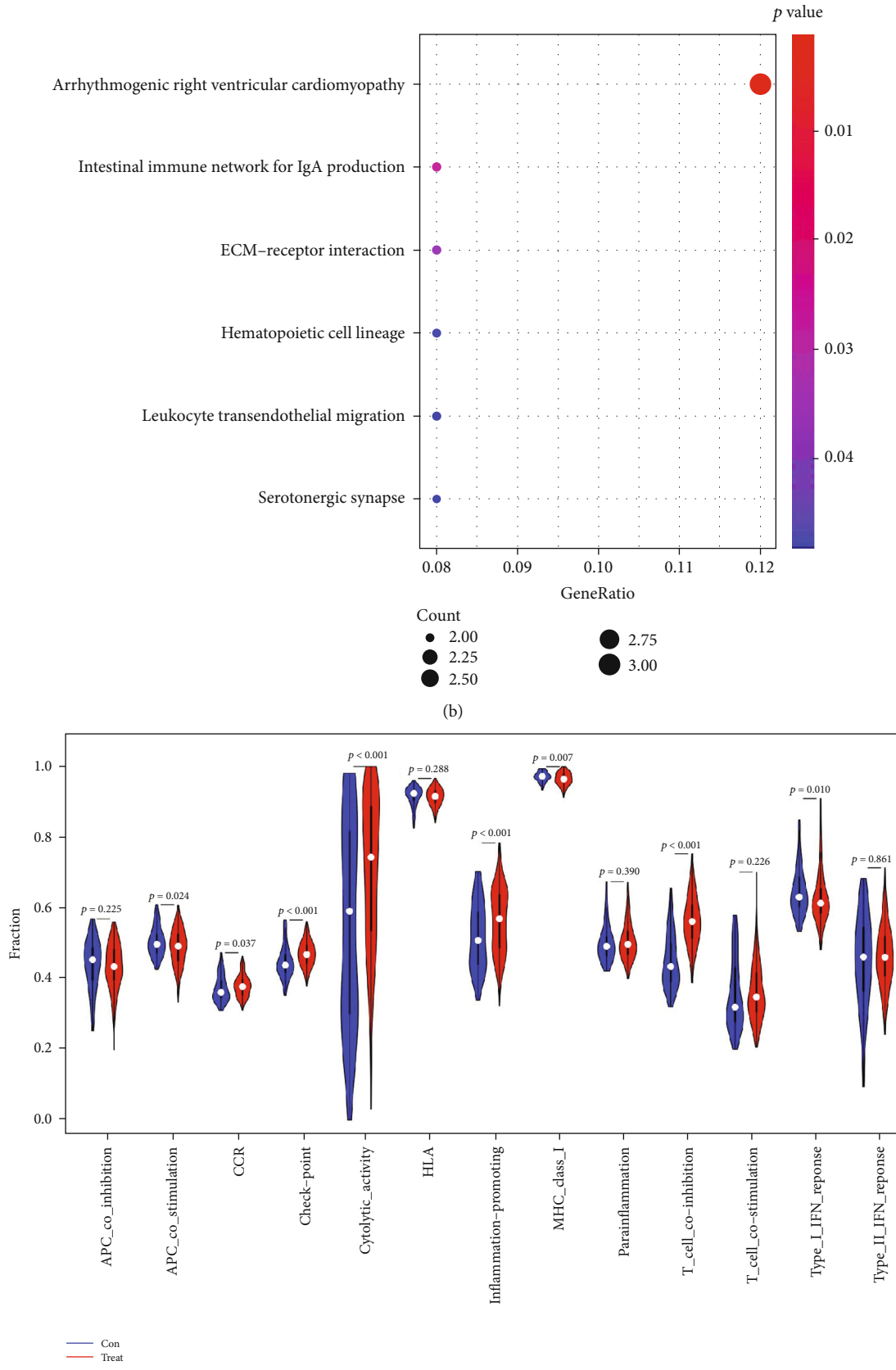
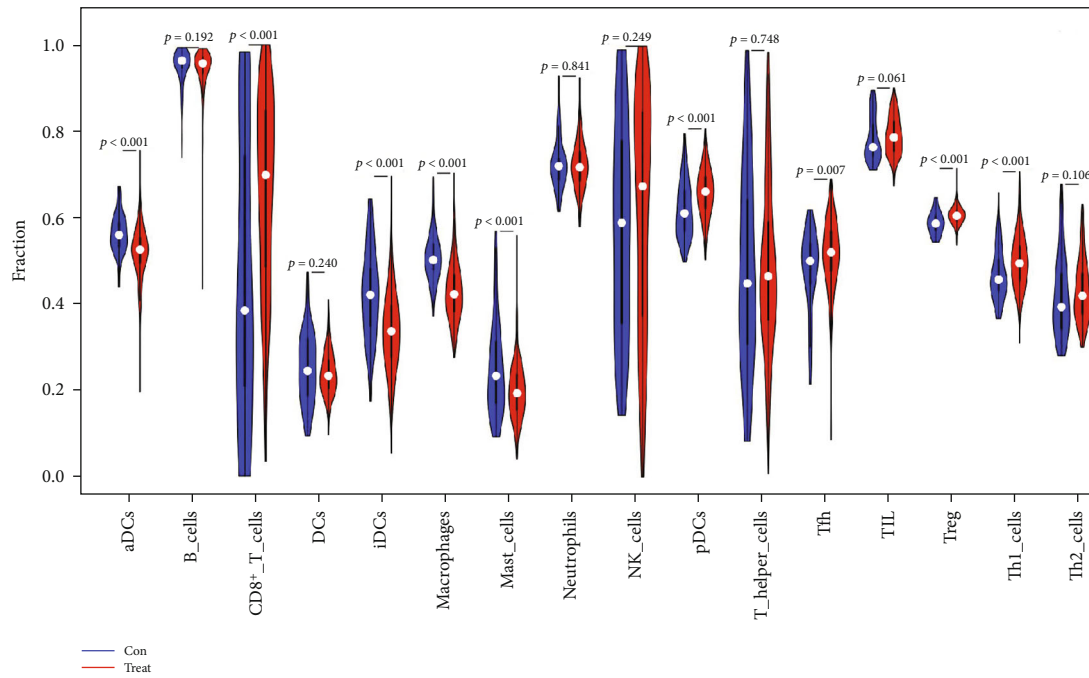
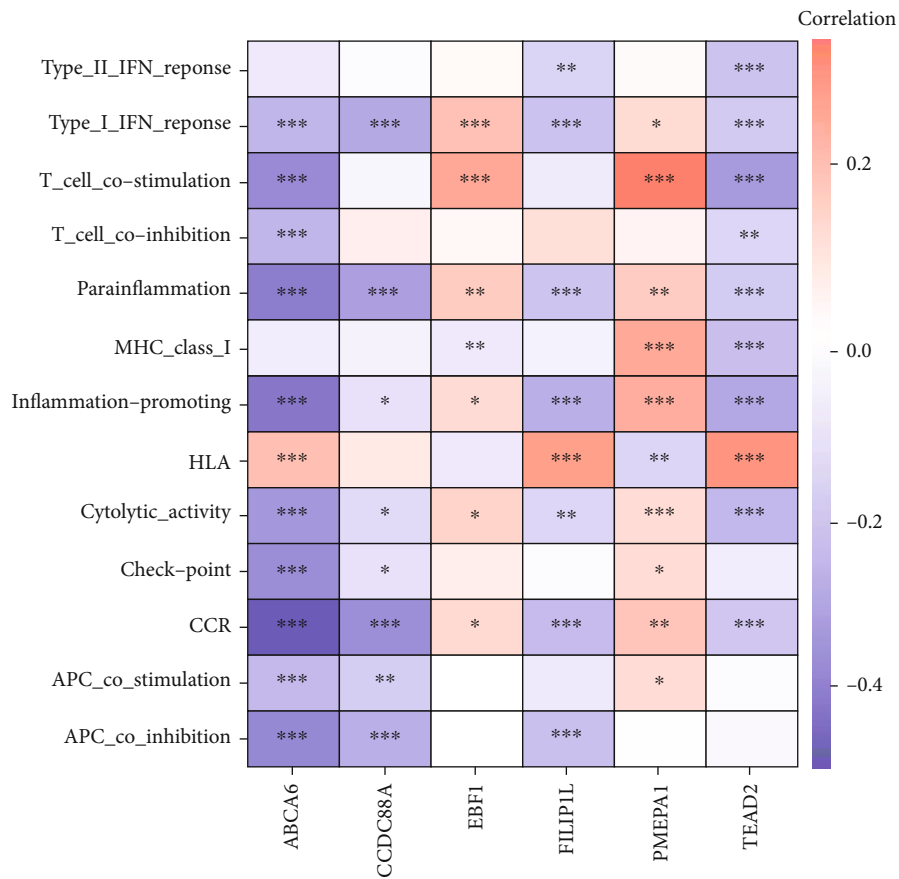


FIGURE 7: Continued.



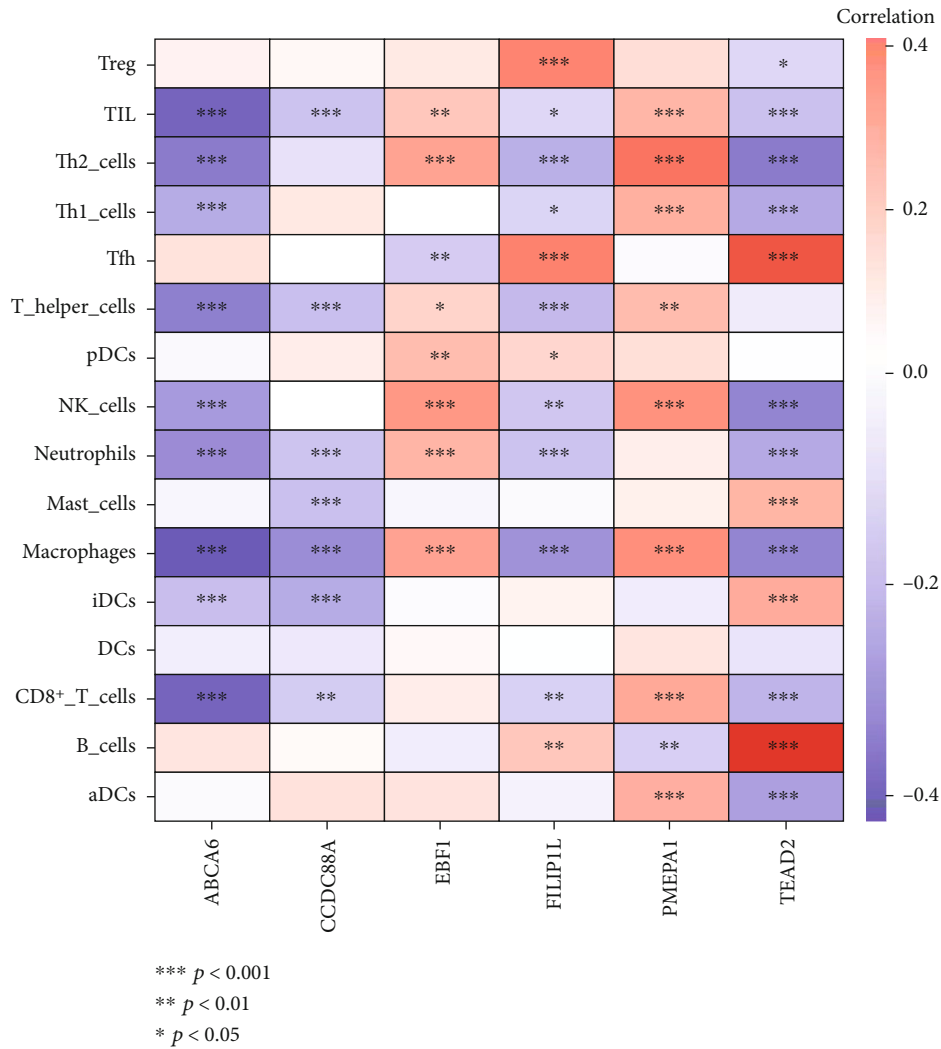
(d)



*** $p < 0.001$
 ** $p < 0.01$
 * $p < 0.05$

(e)

FIGURE 7: Continued.



(f)

FIGURE 7: Functional enrichment analysis. (a) The bubble diagram of the gene ontology (GO) analysis; (b) the bubble diagram of the Kyoto Encyclopedia of Genes and Genome (KEGG) enrichment analysis; (c) the violin plot of the immune-related functions; (d) the violin plot of the immune-related cells; (e) the heatmap of the correlations between the hub genes and the immune-related functions; (f) The heatmap of correlations between the hub genes and the immune-related cells.

This study is not without limitations. First, multiple datasets were retrieved from GEO database, which could have increased the potential heterogeneity due to disparate annotation platforms and clinical covariates of the samples included. The batch effects among datasets could not be completely eliminated. Second, we used public data in our investigation, and thus prospective investigations are needed to validate its predictive power in research with clinical samples. Finally, the underlying molecular mechanisms of the novel biomarkers were not undermined and confirmed by *in vivo* or *in vitro* studies. Hence, the aforementioned deficiencies will be addressed in our further research.

5. Conclusion

In conclusion, six genes (*ABCA6*, *CCDC88A*, *PMEPA1*, *EBF1*, *FILIP1L*, and *TEAD2*) for CLL diagnosis were identified by bioinformatics analysis. In addition, the constructed

ROC curves confirmed that these genes possessed a good diagnostic value for CLL with high sensitivity and accuracy. Integrated analyses revealed significant interactions between these hub genes and tumor immunity, indicating that the biomarkers may promote CLL cell tumorigenesis and survival via the suppressed immune response through its immune-related pathways. Machine learning algorithms identified that these genes could be possible CLL marker genes, providing the foundation for further experimental studies. However, future research is needed to evaluate the performance of these hub genes and their precise underlying mechanism of action in clinical practice.

Data Availability

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Conflicts of Interest

The authors declare no conflict of interest regarding the publication of this paper.

Authors' Contributions

Y. Zhu performed the conception and design of the study, the analysis and interpretation of data, and completed the manuscript. X. Gan conducted the revision of the manuscript. R. Qin and Z. Lin provided bioinformatic guidance. All authors read and gave their approval for publication of the final version of the manuscript.

Acknowledgments

This work was sponsored by the Shanghai Municipal Health and Family Planning Commission (201640038 and 20214Y0375).

References

- [1] M. Hallek, T. D. Shanafelt, and B. Eichhorst, "Chronic lymphocytic leukaemia," *The Lancet*, vol. 391, no. 10129, pp. 1524–1537, 2018.
- [2] National Cancer Institute Surveillance E, and End Results Program, "Cancer Stat Facts: Chronic Lymphocytic Leukemia (CLL)," <https://seer.cancer.gov/statfacts/html/clyl.html>.
- [3] B. Eichhorst, T. Robak, E. Montserrat et al., "Chronic lymphocytic leukaemia: ESMO clinical practice guidelines for diagnosis, treatment and follow-up," *Annals of Oncology: Official Journal of the European Society for Medical Oncology*, vol. 32, no. 1, pp. 23–33, 2021.
- [4] M. Hallek, "Chronic lymphocytic leukemia: 2020 update on diagnosis, risk stratification and treatment," *American Journal of Hematology*, vol. 94, no. 11, pp. 1266–1287, 2019.
- [5] K. R. Rai and P. Jain, "Chronic lymphocytic leukemia (CLL)-then and now," *American Journal of Hematology*, vol. 91, no. 3, pp. 330–340, 2016.
- [6] J. Lee and Y. L. Wang, "Prognostic and predictive molecular biomarkers in chronic lymphocytic leukemia," *The Journal of Molecular Diagnostics: JMD*, vol. 22, no. 9, pp. 1114–1125, 2020.
- [7] A. C. Rawstron, K. A. Kreuzer, A. Soosapilla et al., "Reproducible diagnosis of chronic lymphocytic leukemia by flow cytometry: an European research initiative on CLL (ERIC) & European Society for Clinical Cell Analysis (ESCCA) harmonisation project," *Cytometry Part B, Clinical Cytometry*, vol. 94, no. 1, pp. 121–128, 2018.
- [8] A. Navarro, S. Beà, P. Jares, and E. Campo, "Molecular pathogenesis of mantle cell lymphoma," *Hematology/Oncology Clinics of North America*, vol. 34, no. 5, pp. 795–807, 2020.
- [9] P. Mishra, S. Padhi, P. Ayyanar et al., "Clinicopathological and Immunohistochemical profile of mantle cell lymphoma: an institutional experience," *Cureus*, vol. 13, no. 7, article e16534, 2021.
- [10] T. Yoshino, T. Tanaka, and Y. Sato, "Differential diagnosis of chronic lymphocytic leukemia/small lymphocytic lymphoma and other indolent lymphomas, including mantle cell lymphoma," *Journal of Clinical and Experimental Hematopathology: JCEH*, vol. 60, no. 4, pp. 124–129, 2020.
- [11] W. G. Wierda, J. Brown, J. S. Abramson et al., "NCCN guidelines® insights: chronic lymphocytic leukemia/small lymphocytic lymphoma, version 3.2022," *Journal of the National Comprehensive Cancer Network: JNCCN*, vol. 20, no. 6, pp. 622–634, 2022.
- [12] P. McKay, M. Leach, R. Jackson, G. Cook, S. Rule, and British Committee for Standards in Haematology, "Guidelines for the investigation and management of mantle cell lymphoma," *British Journal of Haematology*, vol. 159, no. 4, pp. 405–426, 2012.
- [13] C. Tarella, L. Arcaini, L. Baldini et al., "Italian Society of Hematology, Italian Society of Experimental Hematology, and Italian Group for Bone Marrow Transplantation guidelines for the management of indolent, nonfollicular B-cell lymphoma (marginal zone, lymphoplasmacytic, and small lymphocytic lymphoma)," *Clinical Lymphoma, Myeloma & Leukemia*, vol. 15, no. 2, pp. 75–85, 2015.
- [14] W. G. Morice, P. J. Kurtin, J. M. Hodnefield et al., "Predictive value of blood and bone marrow flow cytometry in B-cell lymphoma classification: comparative analysis of flow cytometry and tissue biopsy in 252 patients," *Mayo Clinic Proceedings*, vol. 83, no. 7, pp. 776–785, 2008.
- [15] A. Saksena, C. C. Yin, J. Xu et al., "CD23 expression in mantle cell lymphoma is associated with CD200 expression, leukemic non-nodal form, and a better prognosis," *Human Pathology*, vol. 89, pp. 71–80, 2019.
- [16] P. Strati and T. D. Shanafelt, "Monoclonal B-cell lymphocytosis and early-stage chronic lymphocytic leukemia: diagnosis, natural history, and risk stratification," *Blood*, vol. 126, no. 4, pp. 454–462, 2015.
- [17] C. Demir, E. Kara, Ö. Ekinçi, and S. Ebiñç, "Clinical and laboratory features of CD5-negative chronic lymphocytic leukemia," *Medical Science Monitor: International Medical Journal of Experimental and Clinical Research*, vol. 23, pp. 2137–2142, 2017.
- [18] E. L. van Dijk, H. Auger, Y. Jaszczyszyn, and C. Thermes, "Ten years of next-generation sequencing technology," *Trends in Genetics: TIG*, vol. 30, no. 9, pp. 418–426, 2014.
- [19] J. H. Hung and Z. Weng, "Analysis of microarray and RNA-seq expression profiling data," *Cold Spring Harbor Protocols*, vol. 2017, no. 3, p. pdb.top093104, 2017.
- [20] W. Liao, G. Jordaan, P. Nham, R. T. Phan, M. Pelegrini, and S. Sharma, "Gene expression and splicing alterations analyzed by high throughput RNA sequencing of chronic lymphocytic leukemia specimens," *BMC Cancer*, vol. 15, no. 1, p. 714, 2015.
- [21] D. Huang and I. Ovcharenko, "Epigenetic and genetic alterations and their influence on gene regulation in chronic lymphocytic leukemia," *BMC Genomics*, vol. 18, no. 1, p. 236, 2017.
- [22] G. Öztan, M. Aktan, S. Palanduz et al., "Relationship between chromosomal aberrations and gene expressions in the p53 pathway in chronic lymphocytic leukemia," *Balkan Journal of Medical Genetics: BJMG*, vol. 23, no. 1, pp. 15–24, 2020.
- [23] F. Palacios, X. J. Yan, G. Ferrer et al., "Musashi 2 influences chronic lymphocytic leukemia cell survival and growth making it a potential therapeutic target," *Leukemia*, vol. 35, no. 4, pp. 1037–1052, 2021.
- [24] M. Bagheri, B. Khansarinejad, G. Mosayebi, A. Moradabadi, and M. Mondanizadeh, "Alterations in the plasma expression of mir-15b, mir-195 and the tumor-suppressor gene DLEU7 in patients with B-cell chronic lymphocytic leukemia," *Reports*

- of Biochemistry & Molecular Biology*, vol. 10, no. 1, pp. 20–29, 2021.
- [25] P. Clote, “Are RNA networks scale-free?,” *Journal of Mathematical Biology*, vol. 80, no. 5, pp. 1291–1321, 2020.
- [26] N. Lin and H. Zhao, “Are scale-free networks robust to measurement errors?,” *BMC Bioinformatics*, vol. 6, no. 1, p. 119, 2005.
- [27] M. Yang, H. He, T. Peng, Y. Lu, and J. Yu, “Identification of 9 gene signatures by WGCNA to predict prognosis for colon adenocarcinoma,” *Computational Intelligence and Neuroscience*, vol. 2022, Article ID 8598046, 15 pages, 2022.
- [28] J. Li, Y. Qin, and H. Zhang, “Identification of key miRNA-gene pairs in chronic lymphocytic leukemia through integrated analysis of mRNA and miRNA microarray,” *Oncology Letters*, vol. 15, no. 1, pp. 361–367, 2018.
- [29] C. Gao, C. Zhou, J. Zhuang et al., “Identification of key candidate genes and miRNA-mRNA target pairs in chronic lymphocytic leukemia by integrated bioinformatics analysis,” *Molecular Medicine Reports*, vol. 19, no. 1, pp. 362–374, 2019.
- [30] P. Langfelder and S. Horvath, “WGCNA: an R package for weighted correlation network analysis,” *BMC Bioinformatics*, vol. 9, no. 1, p. 559, 2008.
- [31] B. Zhang and S. Horvath, “A general framework for weighted gene co-expression network analysis,” *Statistical Applications in Genetics and Molecular Biology*, vol. 4, p. 17, 2005.
- [32] R. Tibshirani, “The lasso method for variable selection in the cox model,” *Statistics in Medicine*, vol. 16, no. 4, pp. 385–395, 1997.
- [33] D. M. McNeish, “Using lasso for predictor selection and to assuage overfitting: a method long overlooked in behavioral sciences,” *Multivariate Behavioral Research*, vol. 50, no. 5, pp. 471–484, 2015.
- [34] J. Gai, M. Ji, C. Shi et al., “FoxO regulates expression of ABCA6, an intracellular ATP-binding-cassette transporter responsive to cholesterol,” *The International Journal of Biochemistry & Cell Biology*, vol. 45, no. 11, pp. 2651–2659, 2013.
- [35] J. Klucken, C. Buchler, E. Orso et al., “ABCG1 (ABC8), the human homolog of the drosophila white gene, is a regulator of macrophage cholesterol and phospholipid transport,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 97, no. 2, pp. 817–822, 2000.
- [36] X. Bai, L. Jing, Y. Li et al., “TM6PA1 inhibits TGF- β signaling by promoting lysosome degradation of TGF- β receptor and contributes to lung cancer development,” *Cellular Signalling*, vol. 26, no. 9, pp. 2030–2039, 2014.
- [37] N. Nakano, K. Maeyama, N. Sakata et al., “C18 ORF1, a novel negative regulator of transforming growth factor- β signaling,” *The Journal of Biological Chemistry*, vol. 289, no. 18, pp. 12680–12692, 2014.
- [38] C. Y. Yang, S. Ramamoorthy, S. Boller et al., “Interaction of CCR4-NOT with EBF1 regulates gene-specific transcription and mRNA stability in B lymphopoiesis,” *Genes & Development*, vol. 30, no. 20, pp. 2310–2324, 2016.
- [39] R. Gisler, S. E. Jacobsen, and M. Sigvardsson, “Cloning of human early B-cell factor and identification of target genes suggest a conserved role in B-cell development in man and mouse,” *Blood*, vol. 96, no. 4, pp. 1457–1464, 2000.
- [40] H. M. Bond, M. Mesuraca, E. Carbone et al., “Early hematopoietic zinc finger protein (EHZF), the human homolog to mouse Evi3, is highly expressed in primitive human hematopoietic cells,” *Blood*, vol. 103, no. 6, pp. 2062–2070, 2004.
- [41] M. Kwon, E. Hanna, D. Lorang et al., “Functional characterization of filamin a interacting protein 1-like, a novel candidate for antivascular cancer therapy,” *Cancer Research*, vol. 68, no. 18, pp. 7332–7341, 2008.
- [42] B. Zhao, X. Ye, J. Yu et al., “TEAD mediates YAP-dependent gene induction and growth control,” *Genes & Development*, vol. 22, no. 14, pp. 1962–1971, 2008.
- [43] H. Zhang, C. Y. Liu, Z. Y. Zha et al., “TEAD transcription factors mediate the function of TAZ in cell growth and epithelial-mesenchymal transition,” *The Journal of Biological Chemistry*, vol. 284, no. 20, pp. 13355–13362, 2009.
- [44] C. Lin, J. Ear, K. Midde et al., “Structural basis for activation of trimeric Gi proteins by multiple growth factor receptors via GIV/Girdin,” *Molecular Biology of the Cell*, vol. 25, no. 22, pp. 3654–3671, 2014.
- [45] I. López-Sánchez, M. Garcia-Marcos, Y. Mittal, N. Aznar, M. G. Farquhar, and P. Ghosh, “Protein kinase C-theta (PKC θ) phosphorylates and inhibits the guanine exchange factor, GIV/Girdin,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 110, no. 14, pp. 5510–5515, 2013.
- [46] P. Ghosh, A. O. Beas, S. J. Bornheimer et al., “A G {alpha}-GIV molecular complex binds epidermal growth factor receptor and determines whether cells migrate or proliferate,” *Molecular Biology of the Cell*, vol. 21, no. 13, pp. 2338–2354, 2010.
- [47] C. Lin, J. Ear, Y. Pavlova et al., “Tyrosine phosphorylation of the G α -interacting protein GIV promotes activation of phosphoinositide 3-kinase during cell migration,” *Science Signaling*, vol. 4, no. 192, p. ra64, 2011.
- [48] Q. Hu, Y. Li, H. Chen, H. Liao, Y. He, and Q. Zheng, “CCDC88A post-transcriptionally regulates VEGF via miR-101 and subsequently regulates hepatocellular carcinoma,” *Frontiers in Immunology*, vol. 13, p. 859331, 2022.
- [49] W. Liu, S. C. Almo, and X. Zang, “Co-stimulate or co-inhibit regulatory T cells, which side to go?,” *Immunological Investigations*, vol. 45, no. 8, pp. 813–831, 2016.
- [50] L. Chen and D. B. Flies, “Molecular mechanisms of T cell co-stimulation and co-inhibition,” *Nature Reviews Immunology*, vol. 13, no. 4, pp. 227–242, 2013.
- [51] J. C. Riches, J. K. Davies, F. McClanahan et al., “T cells from CLL patients exhibit features of T-cell exhaustion but retain capacity for cytokine production,” *Blood*, vol. 121, no. 9, pp. 1612–1621, 2013.
- [52] G. Görgün, T. A. Holderried, D. Zahrieh, D. Neuberg, and J. G. Gribben, “Chronic lymphocytic leukemia cells induce changes in gene expression of CD4 and CD8 T cells,” *The Journal of Clinical Investigation*, vol. 115, no. 7, pp. 1797–1805, 2005.
- [53] G. D’Arena, L. Laurenti, M. M. Minervini et al., “Regulatory T-cell number is increased in chronic lymphocytic leukemia patients and correlates with progressive disease,” *Leukemia Research*, vol. 35, no. 3, pp. 363–368, 2011.
- [54] B. Pourghesari, R. Bruton, H. Parry et al., “The number of cytomegalovirus-specific CD4+ T cells is markedly expanded in patients with B-cell chronic lymphocytic leukemia and determines the total CD4+ T-cell repertoire,” *Blood*, vol. 116, no. 16, pp. 2968–2974, 2010.
- [55] M. Beyer, M. Kochanek, K. Darabi et al., “Reduced frequencies and suppressive function of CD4+CD25hi regulatory T cells in patients with chronic lymphocytic leukemia after therapy with fludarabine,” *Blood*, vol. 106, no. 6, pp. 2018–2025, 2005.

- [56] V. Sasidharan Nair and E. Elkord, "Immune checkpoint inhibitors in cancer therapy: a focus on T-regulatory cells," *Immunology and Cell Biology*, vol. 96, no. 1, pp. 21–33, 2018.
- [57] S. Sakaguchi, N. Mikami, J. B. Wing, A. Tanaka, K. Ichiyama, and N. Ohkura, "Regulatory T cells and human disease," *Annual Review of Immunology*, vol. 38, no. 1, pp. 541–566, 2020.
- [58] A. Tanaka and S. Sakaguchi, "Targeting Treg cells in cancer immunotherapy," *European Journal of Immunology*, vol. 49, no. 8, pp. 1140–1146, 2019.
- [59] N. Chihara, A. Madi, T. Kondo et al., "Induction and transcriptional regulation of the co-inhibitory gene module in T cells," *Nature*, vol. 558, no. 7710, pp. 454–459, 2018.
- [60] L. E. Lucca and M. Dominguez-Villar, "Modulation of regulatory T cell function and stability by co-inhibitory receptors," *Nature Reviews Immunology*, vol. 20, no. 11, pp. 680–693, 2020.
- [61] V. Griggio, F. Perutelli, C. Salvetti et al., "Immune dysfunctions and immune-based therapeutic interventions in chronic lymphocytic leukemia," *Frontiers in Immunology*, vol. 11, p. 594556, 2020.