Oncology Treatment Discovery, 2024, Volume 2, Issue 4
http://ojs.bbwpublisher.com/index.php/OTD
Online ISSN: 2981-8079

Identification of Immune-Related Feature Genes in Ovarian Cancer Using Bioinformatics and Analysis of Immune Cell Infiltration

Huilu Cao, Yidong Fan, Xiaolin Zhang, Jing Wang, Zheng Huang, Jiongzuo Pan, Shujuan Fan, Qian Tan*

Guangxi University of Chinese Medicine, Nanning 530200, Guangxi Province, China

***Corresponding author:** Qian Tan, 544804425@qq.com

Abstract: *Objective:* To identify immune-related feature genes in ovarian cancer through bioinformatics analysis and perform immune-related investigations, which hold significant value for the early diagnosis and prevention of ovarian cancer. *Methods:* Bioinformatics analysis was utilized to identify immune-related feature genes in ovarian cancer. The GSE18520 and GSE40595 datasets were downloaded from the Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/geo/) based on the gene expression comprehensive database, and the corresponding platform's chip probe information was retrieved. GSE18520 served as the training set, and GSE40595 served as the validation set. A total of 2660 immune response genes (IRGs) were obtained from the ImmPort database (https://www.immport.org/home). Immune genes were screened and analyzed for feature genes using the "limma" package of R (4.2.1) software, and the results were visualized in a heat map. LASSO regression analysis and ssGSEA analysis were conducted to investigate the distribution of immune cell infiltration. Changes in regression coefficients of different genes in the model were also analyzed. *Results:* Five key genes—*CLEC4M*, *DEFB1*, *LCN2*, *PTH2R*, and *LGALS2*—were identified, and the correlation between these key genes and immune cells was analyzed. *Conclusion:* The findings indicate that *CLEC4M*, *DEFB1*, *LCN2*, *PTH2R*, and *LGALS2* are significantly associated with various immune cell types, suggesting that these genes may regulate immune cell behavior and influence disease progression. This bioinformatics study provides a foundation for potential therapeutic targets in ovarian cancer; however, further clinical and experimental studies are required to validate the findings.

Keywords: Ovarian cancer; Immune cells; Bioinformatics; Genes

Online publication: January 3, 2025

1. Introduction

Ovarian cancer (OC) is one of the leading causes of death among gynecologic malignancies [1]. Due to the lack of distinct early symptoms, OC is often diagnosed at advanced stages. Reliable diagnostic markers and early detection methods remain insufficient, emphasizing the need to improve early recognition of OC among health professionals

and the general female population [2]. Therefore, identifying early warning indicators for OC prognosis and further exploring its molecular mechanisms can provide critical theoretical guidance for early intervention [3].

Over the past decade, progress in chemotherapy has been slow in improving the prognosis of OC, prompting increased research into molecular-targeted therapies. Similar to other cancer types, OC exhibits significant heterogeneity across different subtypes and individual tumors, posing major challenges to the effectiveness of targeted drug therapies [4]. This heterogeneity, a hallmark of many cancers including OC, has potential predictive value for survival outcomes following chemotherapy, particularly in high-grade serous ovarian cancer [5].

In China, the prevalence of OC has shown a significant upward trend over the past 30 years, with a notable acceleration in the last five years. The most affected population consists of women over 40 years of age, particularly postmenopausal and elderly women [6]. It is projected that the number of OC patients in China will continue to rise at a rate surpassing the global average over the next decade. Currently, surgical intervention remains the primary treatment for OC, serving to confirm tumor type and disease stage [7]. Postoperative chemotherapy with platinum and taxane drugs is then administered. In the past two decades, the rapid development of immunotherapy has introduced new possibilities for OC treatment, with the screening of potential immunotherapeutic drugs and the provision of adjuvant immunotherapy offering the potential to extend patient survival [8].

Bioinformatics, an interdisciplinary field combining biology, information science, and statistics, plays a critical role in cancer research [9]. It encompasses the processing and quality control of raw sequencing data, variant detection and annotation, integration of diverse molecular data types, visualization, and the generation of interpretable data reports. In recent years, bioinformatics has been widely applied in cancer diagnosis and treatment, driven by the need to convert biological data into actionable knowledge [10]. The integration and analysis of bioinformatics-generated big data are essential for personalized healthcare and genomics, establishing bioinformatics as a cornerstone of precision oncology [11].

Given the advancements in bioinformatics, future research on immune-related genes in OC is of great importance. In this study, the ovarian cancer dataset (accession number GSE18520) was retrieved from the Gene Expression Omnibus (GEO) database. Immune-related genes were screened, and characteristic genes were analyzed using R software packages such as “limma,” with the aim of elucidating the pathogenesis of OC and providing new references for early diagnosis and treatment strategies.

2. Materials and methods

2.1. Data download and processing

Microarray or high-throughput data were downloaded from the Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/geo/) using “ovarian cancer” as the keyword, with the species set to “*Homo sapiens*.” The dataset was required to include both normal and disease samples. Datasets GSE18520 and GSE40595, along with their corresponding platform chip probe information, were retrieved and downloaded. GSE18520 was designated as the training group, while GSE40595 served as the validation group. During the conversion of probe IDs to gene symbols, if multiple probes corresponded to a single gene symbol, the average expression level was calculated to represent the gene expression level. This conversion was performed using Perl (version 5.10.1). Additionally, 2660 immune response genes (IRGs) were obtained from the ImmPort database (https://www.immport.org/home).

2.2. Differential expression analysis

Differential expression analysis was conducted using the "limma" package in R (version 4.2.1). Differentially expressed genes (DEGs) were identified based on an adjusted *P*-value < 0.05 and an absolute value of the $\log_2$ fold change ≥ 1 ($|\log_2 FC| \geq 1$). Heatmap visualization of DEGs was performed using the "pheatmap" package.

2.3. Screening of immune-related feature genes in ovarian cancer

Two machine learning algorithms, Support Vector Machine Recursive Feature Elimination (SVM-RFE) and Least Absolute Shrinkage and Selection Operator (LASSO), were employed to predict immune-related feature genes in OC. The Lasso regression algorithm was constructed using the "glmnet" package, with ten-fold cross-validation to determine the optimal λ value. L1 regularization was applied to enhance prediction accuracy and facilitate feature gene selection. SVM-RFE, a supervised learning algorithm commonly used for classification and regression analysis, was used to score genes and iteratively select those with strong classification performance. The "e1071" package was utilized to implement the SVM-RFE algorithm.

2.4. Validation of feature genes and evaluation of diagnostic accuracy

The expression levels of feature genes were validated using the GSE40595 dataset obtained from the GEO database. The "limma" package was used to extract the expression levels of feature genes, and differential analysis was performed with thresholds of $|\log_2 FC| \geq 1$ and adj. *P*-value < 0.05. Violin plots of differentially expressed genes were generated using the "ggunchained" package. The diagnostic performance of each feature gene was evaluated by plotting receiver operating characteristic (ROC) curves using the "pROC" package.

2.5. Immune cell infiltration analysis

Single-sample gene set enrichment analysis (ssGSEA) was employed to compare gene expression data from samples with immune cell gene sets, calculating the relative abundance of immune cells in each sample. The ssGSEA algorithm was used to estimate the infiltration abundance of immune cells in normal and ovarian cancer samples, and immune cell differences between the two groups were visualized. Correlation analysis among immune genes, immune cells, and feature genes was conducted using the "ggcorrplot" package. A correlated network heatmap was generated using the "linkET" package.

2.6. Construction of the ceRNA network for feature genes

miRNAs associated with feature genes were predicted using the miRDB (https://mirdb.org/), TargetScan (http://www.targetscan.org/vert_72/), and miRTarBase (http://mirtarbase.mbc.nctu.edu.tw/) databases. The intersection of prediction results from these three miRNA databases was identified as the target genes of feature gene-related miRNAs. Long non-coding RNAs (lncRNAs) targeting key miRNAs were predicted using the spongeScan (http://spongescan.rc.ufl.edu/) database. A ceRNA network of mRNA-miRNA-lncRNA for feature genes was subsequently constructed.

2.7. Statistical analysis

All bioinformatics analyses were performed using R software (version 4.1.2, 64-bit). Independent sample *t*-tests were conducted for comparisons between the two groups.

3. Results

3.1. Differential expression analysis

A differential expression gene analysis was conducted to identify genes significantly differentially expressed between disease and normal groups. Using a *P*-value threshold of less than 0.05 and an absolute log fold change greater than 1 as screening criteria, 483 key genes were identified. These genes exhibited significant expression differences between the groups, indicating their potential roles in the disease's onset and progression. To visualize the expression patterns of these key genes, a heatmap was generated, illustrating their expression differences across samples (see **Figure 1**). The heatmap revealed distinct expression patterns between the disease and normal groups, establishing a foundation for subsequent in-depth analyses.

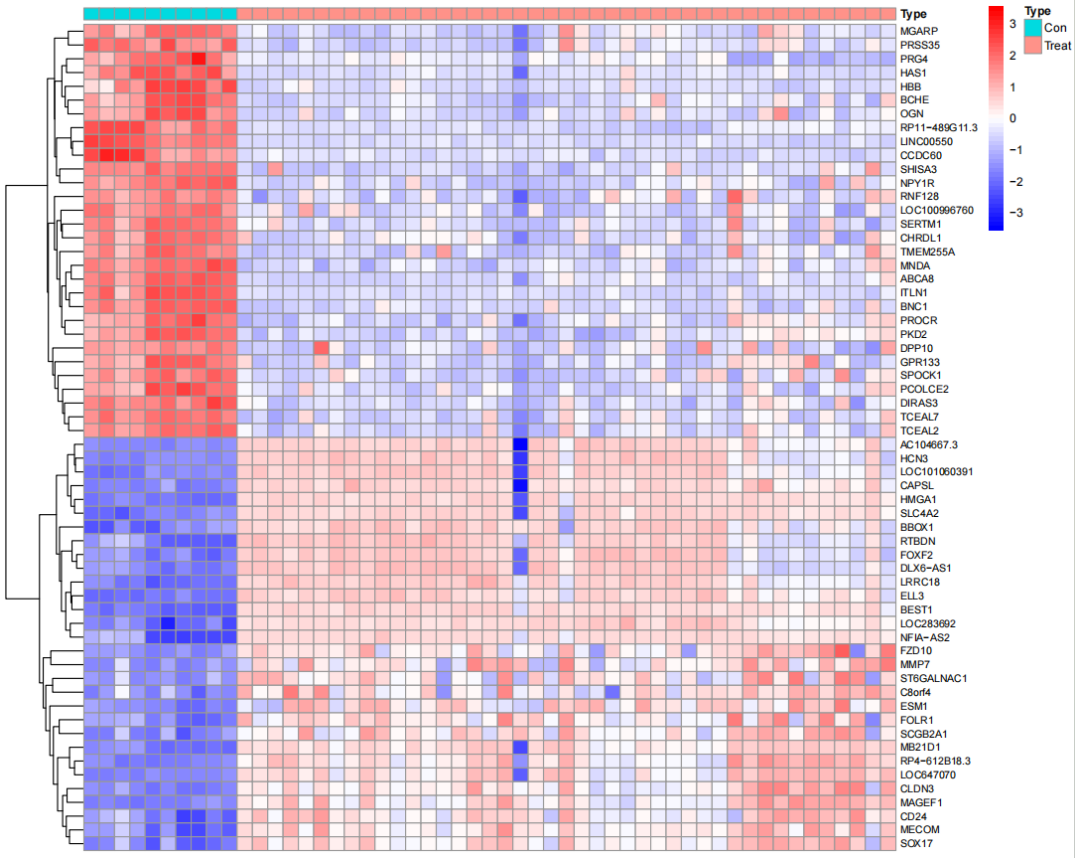

Figure 1. Heatmap of key genes. This heatmap displays the expression profiles of the 483 key genes with significant differences between the disease and normal groups. Each row represents a gene, and each column represents a sample. Red indicates gene upregulation, while blue indicates downregulation. The distinct expression patterns highlight the potential involvement of these genes in disease progression.

3.2. Key gene selection

To identify core disease-related genes among the DEGs, an intersection analysis was performed between the DEGs and known immune-related feature genes. This analysis identified 51 genes implicated in immune-related pathways (**Figure 2**). Further screening using the LASSO regression model narrowed the selection to five key genes: *CLEC4M*, *KCNH2*, *AKT3*, *TNFRSF8*, and *FCN1* (see **Figures 3** and **4**). These genes are hypothesized to play critical roles in immune regulation and disease progression, forming the focus of subsequent investigations.

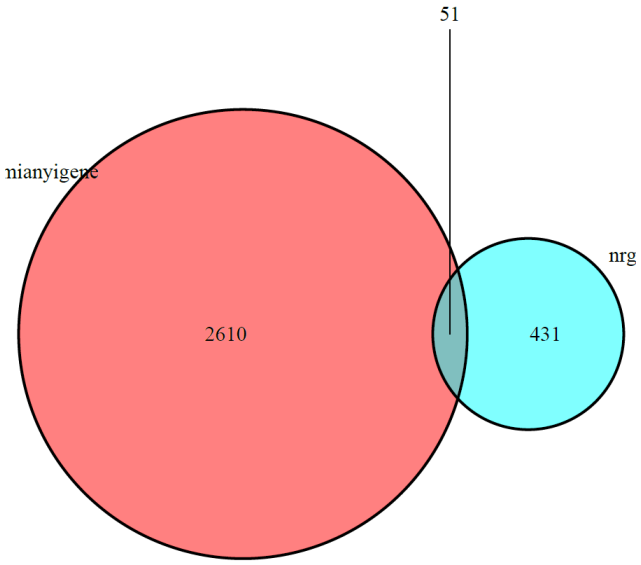

Figure 2. Intersection genes. The left side shows differentially expressed genes, while the right side displays known immune-related characteristic genes. An intersection analysis identified 51 key genes linked to immune-related pathways.

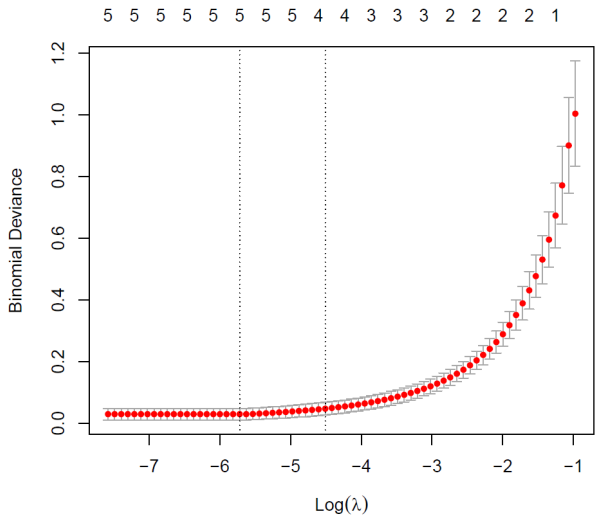

Figure 3. LASSO regression model selection parameter curve. This figure illustrates the relationship between the penalty coefficient (λ) and model performance in the LASSO regression analysis. The x-axis represents logarithmic λ values, and the y-axis shows the binomial deviance. Red dots indicate model deviations at different λ values, while the vertical dashed line identifies the optimal λ for gene selection.

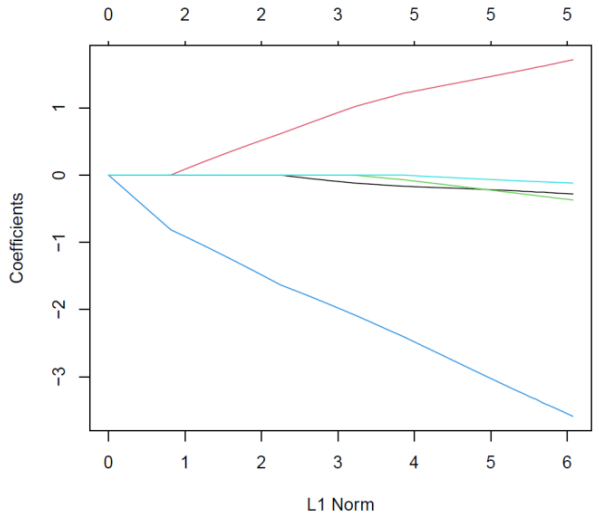

Figure 4. LASSO regression coefficient path diagram. This diagram depicts the changes in regression coefficients of genes as the penalty parameter (λ) varies. The x-axis represents the L1 norm (λ values), and the y-axis represents regression coefficients. Different colored lines correspond to different genes. The five identified key genes are those whose coefficients remain significant as λ increases.

3.3. Immune correlation analysis

To evaluate the relationship between the selected key genes and immune cells, ssGSEA was applied to assess immune cell infiltration. The analysis revealed significant differences in the infiltration levels of multiple immune cell types between the normal and disease groups. For instance, activated CD4 T cells and mast cells showed notable variations in infiltration levels (see **Figure 5**), suggesting their involvement in disease onset and progression.

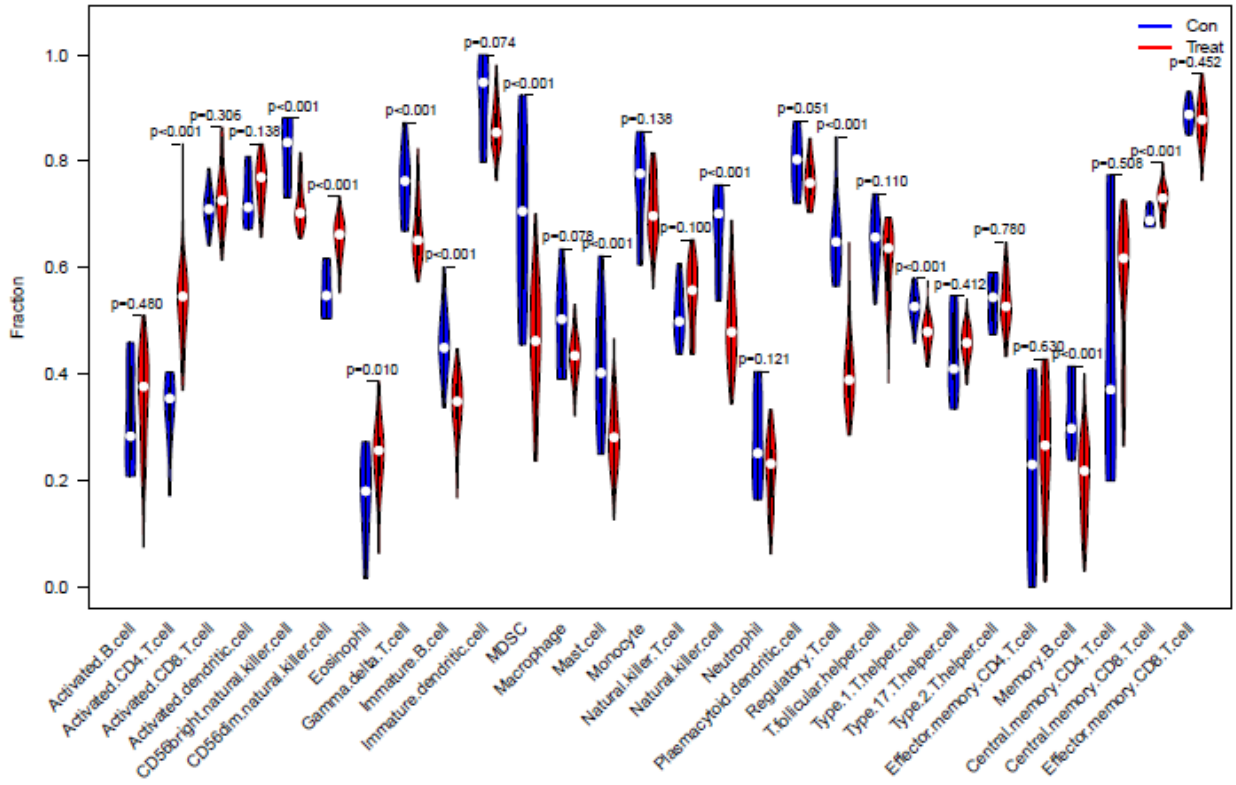

Figure 5. Immune cell infiltration distribution (ssGSEA analysis). This figure illustrates the infiltration levels of various immune cell types between the disease and normal groups. The x-axis represents immune cell types, and the y-axis denotes infiltration proportions. The *P*-values above each cell type indicate the significance of the differences.

Correlation analysis was also conducted to investigate the associations between the five key genes and various immune cell types (see **Figure 6**). The results highlighted a strong correlation between the KCNH2 gene and several immune cell types, particularly immature dendritic cells and natural killer cells. These findings suggest that the selected key genes may influence disease progression through the modulation of immune cell behavior. This analysis provides valuable insights into the immune mechanisms underlying the disease and highlights potential diagnostic and therapeutic targets.

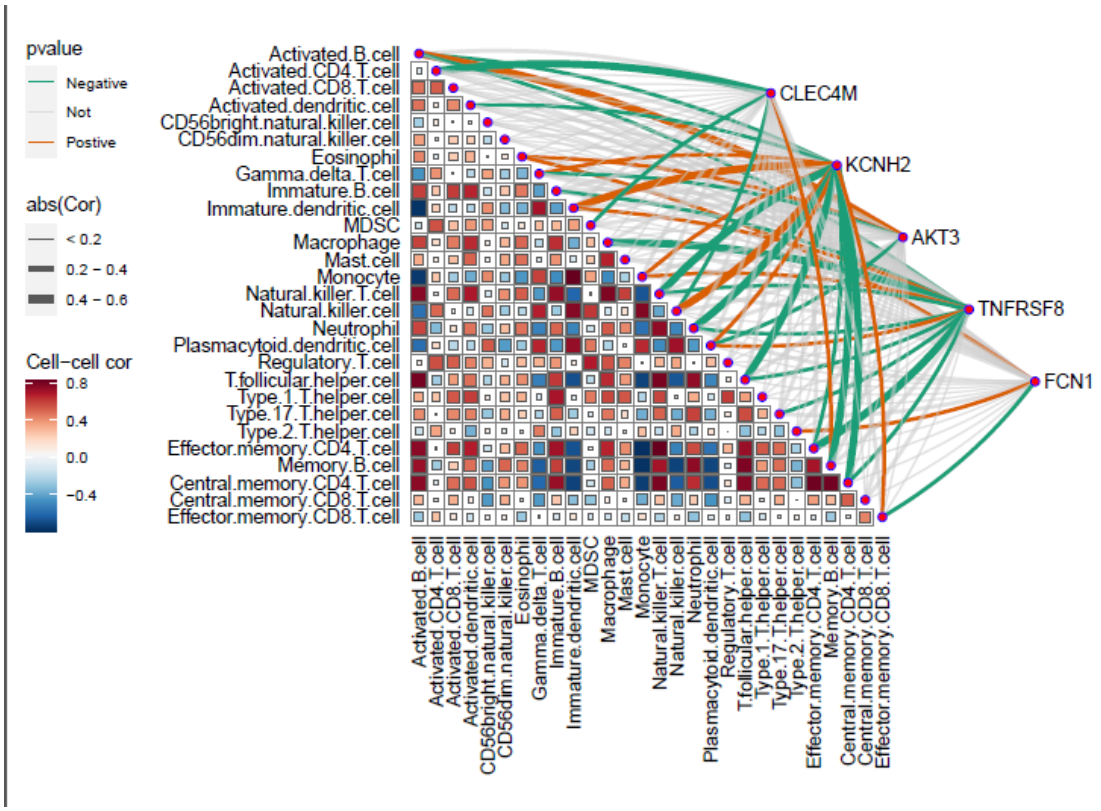

Figure 6. Correlation analysis between key genes and immune cell types. This figure shows the correlation between the five key genes and immune cell types. Line colors represent the correlation direction (orange for positive, green for negative), and line thickness indicates correlation strength. The KCNH2 gene exhibited significant correlations with multiple immune cell types, emphasizing its potential role in modulating immune cell behavior.

4. Discussion

Increasing evidence highlights the critical role of interactions between tumor cells and the tumor microenvironment, particularly the immune microenvironment, in tumor progression [12]. This study utilized bioinformatics to analyze 2,660 immune-related genes from the GSE18520 dataset. Several genes, including *C-type lectin domain family 4, member M* (*CLEC4M*), *beta-defensin 1* (*DEFB1*), *lipocalin 2* (*LCN2*), *parathyroid hormone 2 receptor* (*PTH2R*), and *lectin, galactoside-binding, soluble, 2* (*LGALS2*), were identified as being associated with OC.

CLEC4M was found to potentially play a significant role in the activation and response of immune cells such as T cells, lymphocytes, myeloid leukocytes, and macrophages, reinforcing its involvement in immune activity and its influence on the tumor microenvironment. Studies have demonstrated a correlation between *CLEC4M* and tumor progression [13,14]. Analysis using the KMplot™ database has shown that *CLEC4M* overexpression is linked to recurrence-free, progression-free, and disease-specific survival in patients. *CLEC4M* overexpression inhibits the proliferation of Huh7 and PLC/PRF/5 cells while enhancing apoptosis by suppressing the Janus kinase 1/signal transducer and activator of the transcription 3 pathway, which is implicated in various tumor types [15].

The beta-defensin family is integral to host defense against viral infections, with *DEFB1* recognized as a critical antimicrobial peptide in epithelial cells [16]. Although *DEFB1* is known as a tumor suppressor gene, its

role in OC has not been previously reported [17]. This study suggests that further investigation into *DEFB1* could provide valuable insights and novel approaches for OC treatment.

LCN2, a member of the adipokine protein family, regulates processes associated with cancer cachexia in diseases such as lung, pancreatic, breast, and oral squamous cell carcinomas [18]. It has garnered attention as a therapeutic target for cancers, including OC, where its transcriptional activity may influence cancer cell invasiveness and angiogenic capacity [19,20]. Serum *LCN2* levels have shown potential as biomarkers for epithelial ovarian cancer, warranting further research to determine its utility in early diagnosis and improved sensitivity and specificity for identifying subgroups of OC [21–23].

The parathyroid hormone 2 receptor (PTH2R), concentrated in the endocrine and limbic regions of the forebrain, is a B1 G protein-coupled receptor involved in calcium transport, nociception, and wound healing [24,25]. Analysis has linked PTH2R with the tumor suppressor gene *MUM1L1*, with higher expression levels observed in normal ovarian tissue compared to OC tissue [26]. Reducing PTH2R expression has been shown to inhibit OC growth, invasion, and metastasis [27]. Furthermore, PTH2R expression correlates with increased tumor mutational burden (TMB), suggesting its potential as a future biomarker for OC.

LGALS2, a member of the galactoside-binding galectin family, is associated with immune evasion and disease pathogenesis in conditions such as inflammatory bowel disease, coronary artery disease, and cancer [28]. *LGALS2* increases tumor-associated macrophages, and its inhibition via antibodies has been shown to reverse immunosuppression and prevent tumor growth [29–31]. Elevated *LGALS2* expression is linked to favorable prognoses in diseases like rheumatoid arthritis and thyroid and colorectal cancers [30,33]. However, its role in OC remains unclear and requires further investigation to elucidate its therapeutic potential.

5. Conclusion

In conclusion, this study identified immune-specific targets for OC through systematic and effective screening methods. The findings provide a theoretical framework for further understanding OC pathogenesis and propose new targets for clinical treatment. However, these results are derived from online databases, necessitating additional validation through animal and clinical studies. Such efforts could enhance the clinical diagnosis and treatment of OC and support the development of novel therapeutic agents.

Funding

Guangxi Higher Education Key Laboratory for the Research of Du-related Diseases in Zhuang Medicine (Gui Jiao Ke Yan (2022) No.10); ATCM's Project of High-Level Construction of Key TCM Disciplines/Medicine for Ethnic Minorities (Zhuang Medicine) (Project No.: zyyzdxk-2023164); Natural Science Foundation of Guangxi Province (Project No.: 2022JJB140440)

Disclosure statement

The authors declare no conflict of interest.

References

[1] Kujawa KA, Lisowska KM, 2015, Rak Jajnika – Od Biologii do Kliniki [Ovarian Cancer – From Biology to Clinic]. Postępy Higieny i Medycyny Doświadczalnej, 69: 1275–1290.

[2] Rooth C, 2013, Ovarian Cancer: Risk Factors, Treatment and Management. British Journal of Nursing, 22(17): S23–30.

[3] Ledermann JA, 2016, PARP Inhibitors in Ovarian Cancer. Annals of Oncology, 27(Suppl 1): i40–i44.

[4] Kossaï M, Leary A, Scoazec JY, et al., 2018, Ovarian Cancer: A Heterogeneous Disease. Pathobiology, 85(1–2): 41–49.

[5] Schwarz RF, Ng CK, Cooke SL, et al., 2015, Spatial and Temporal Heterogeneity in High-Grade Serous Ovarian Cancer: A Phylogenetic Analysis. PLoS Medicine, 12(2): e1001789.

[6] Wang Y, Wang Z, Zhang Z, et al., 2023, Burden of Ovarian Cancer in China from 1990 to 2030: A Systematic Analysis and Comparison with the Global Level. Frontiers in Public Health, 11: 1136596.

[7] Cortez AJ, Tudrej P, Kujawa KA, et al., 2018, Advances in Ovarian Cancer Therapy. Cancer Chemother Pharmacol, 81(1): 17–38.

[8] Yang C, Xia BR, Zhang ZC, et al., 2020, Immunotherapy for Ovarian Cancer: Adjuvant, Combination, and Neoadjuvant. Frontiers in Immunology, 11: 577869.

[9] Jäger N, 2022, Bioinformatics Workflows for Clinical Applications in Precision Oncology. Seminars in Cancer Biology, 84: 103–112.

[10] Canzoneri R, Lacunza E, Abba MC, 2019, Genomics and Bioinformatics as Pillars of Precision Medicine in Oncology. Medicina (B Aires), 79(Spec 6/1): 587–592.

[11] Liang Y, Lin F, Huang Y, 2022, Identification of Biomarkers Associated with Diagnosis of Osteoarthritis Patients Based on Bioinformatics and Machine Learning. Journal of Immunology Research, 2022: 5600190.

[12] Wu X, Qin K, Iroegbu CD, et al., 2022, Genetic Analysis of Potential Biomarkers and Therapeutic Targets in Ferroptosis from Coronary Artery Disease. Journal of Cell Molecular Medicine, 26(8): 2177–2190.

[13] Huang J, Zhang J, Wang F, et al., 2022, Comprehensive Analysis of Cuproptosis-Related Genes in Immune Infiltration and Diagnosis in Ulcerative Colitis. Frontiers in Immunology, 13: 1008146.

[14] Liu H, Yu Z, Liu Y, et al., 2023, Investigation of Diagnostic and Prognostic Value of CLEC4M of Non-Small Cell Lung Carcinoma Associated with Immune Microenvironment. International Journal of General Medicine, 16: 1317–1332.

[15] Yu Q, Gao K, 2020, CLEC4M Overexpression Inhibits Progression and is Associated with a Favorable Prognosis in Hepatocellular Carcinoma. Molecular Medicine Reports, 22(3): 2245–2252.

[16] Ling YM, Chen JY, Guo L, et al., 2017, β-Defensin 1 Expression in HCV Infected Liver/Liver Cancer: An Important Role in Protecting HCV Progression and Liver Cancer Development. Scientific Reports, 7(1): 13404.

[17] Álvarez ÁH, Velázquez MM, Montes EPO, 2018, Human β-Defensin 1 Update: Potential Clinical Applications of the Restless Warrior. International Journal of Biochemistry and Cell Biology, 104: 133–137.

[18] Wang D, Li X, Jiao D, et al., 2023, LCN2 Secreted by Tissue-Infiltrating Neutrophils Induces the Ferroptosis and Wasting of Adipose and Muscle Tissues in Lung Cancer Cachexia. Journal of Hematology Oncology, 16(1): 30.

[19] Lemecha M, Chalise JP, Takamuku Y, et al., 2022, Lcn2 Mediates Adipocyte-Muscle-Tumor Communication and Hypothermia in Pancreatic Cancer Cachexia. Molecular Metabolism, 66: 101612.

[20] Bao Y, Yan Z, Shi N, et al., 2024, LCN2: Versatile Players in Breast Cancer. Biomedicine & Pharmacotherapy, 171: 116091.

[21] Huang Z, Rui X, Yi C, et al., 2023, Silencing LCN2 Suppresses Oral Squamous Cell Carcinoma Progression by Reducing EGFR Signal Activation and Recycling. Journal of Experimental & Clinical Cancer Research, 42(1): 60.

[22] Zhao H, Ding F, Zheng G, 2020, LncRNA TMPO-AS1 Promotes LCN2 Transcriptional Activity and Exerts Oncogenic Functions in Ovarian Cancer. FASEB Journal, 34(9): 11382–11394.

[23] Cho H, Kim JH, 2009, Lipocalin2 Expressions Correlate Significantly with Tumor Differentiation in Epithelial Ovarian Cancer. Journal of Histochemistry and Cytochemistry, 57(5): 513–521.

[24] Dobolyi A, Dimitrov E, Palkovits M, et al., 2012, The Neuroendocrine Functions of the Parathyroid Hormone 2 Receptor. Frontiers in Endocrinology (Lausanne), 3: 121.

[25] Wang X, Cheng X, Zhao L, et al., 2021, Molecular Insights into Differentiated Ligand Recognition of the Human Parathyroid Hormone Receptor 2. Proceedings of the National Academy of Sciences of the United States of America, 118(32): e2101279118.

[26] Zhang L, Wu X, Fan X, et al., 2023, MUM1L1 as a Tumor Suppressor and Potential Biomarker in Ovarian Cancer: Evidence from Bioinformatics Analysis and Basic Experiments. Combinatorial Chemistry & High Throughput Screening, 26(14): 2487–2501.

[27] Xiaowei W, Tong L, Yanjun Q, et al., 2022, PTH2R is Related to Cell Proliferation and Migration in Ovarian Cancer: A Multi-Omics Analysis of Bioinformatics and Experiments. Cancer Cell International, 22(1): 148.

[28] Negedu MN, Duckworth CA, Yu LG, 2022, Galectin-2 in Health and Diseases. International Journal of Molecular Sciences, 24(1): 341.

[29] Ji P, Gong Y, Jin ML, et al., 2022, In Vivo Multidimensional CRISPR Screens Identify Lgals2 as an Immunotherapy Target in Triple-Negative Breast Cancer. Science Advances, 8(26): eabl8247.

[30] Chetry M, Bhandari A, Feng R, et al., 2022, Overexpression of Galectin2 (LGALS2) Predicts a Better Prognosis in Human Breast Cancer. American Journal of Translational Research, 14(4): 2301–2316.

[31] Li H, Yu L, Zhang X, et al., 2022, Exploring the Molecular Mechanisms and Shared Gene Signatures between Rheumatoid Arthritis and Diffuse Large B Cell Lymphoma. Frontiers in Immunology, 13: 1036239.

[32] Xu D, Guo L, Zhang S, et al., 2022, LGALS2 Suppresses the Progression of Papillary Thyroid Carcinoma by Regulating the PI3K/AKT Pathway. Gland Surgery, 11(9): 1518–1528.

[33] Li H, Zhao L, Lau YS, et al., 2021, Genome-Wide CRISPR Screen Identifies LGALS2 as an Oxidative Stress-Responsive Gene with an Inhibitory Function on Colon Tumor Growth. Oncogene, 40(1): 177–188.