



OPEN ACCESS

EDITED BY

Domenico Mallardo,
G. Pascale National Cancer Institute Foundation
(IRCCS), Italy

REVIEWED BY

Yunxian Dong,
Shandong First Medical University and
Shandong Academy of Medical Sciences, China
Jin Zhang,
University of Mississippi Medical Center,
United States
Edward Vitkin,
Tel Aviv University, Israel

*CORRESPONDENCE

Yousheng Mao,
✉ maoyousheng1992@163.com

[†]These authors share first authorship

RECEIVED 06 August 2024

ACCEPTED 31 March 2025

PUBLISHED 14 April 2025

CITATION

Xiao Q, Xu P, Xu W, Song Q and Mao Y (2025)
Identification of characteristic genes in
cutaneous squamous cell carcinoma based on
weighted gene co-expression network analysis.
Front. Genet. 16:1470584.
doi: 10.3389/fgene.2025.1470584

COPYRIGHT

© 2025 Xiao, Xu, Xu, Song and Mao. This is an
open-access article distributed under the terms
of the [Creative Commons Attribution License](#)
(CC BY). The use, distribution or reproduction in
other forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in this
journal is cited, in accordance with accepted
academic practice. No use, distribution or
reproduction is permitted which does not
comply with these terms.

Identification of characteristic genes in cutaneous squamous cell carcinoma based on weighted gene co-expression network analysis

Qipeng Xiao^{1†}, Pengfei Xu^{1†}, Wenjun Xu², Qiuhe Song¹ and
Yousheng Mao^{1,3*}

¹Department of Dermatology, Affiliated Hospital of Jiujiang University, Jiujiang, Jiangxi, China, ²College of Nursing, Jiujiang Vocational University, Jiujiang, Jiangxi, China, ³Jiujiang Clinical Precision Medicine Research Center, Affiliated Hospital of Jiujiang University, Jiujiang, Jiangxi, China

Objective: This study aims to identify characteristic genes associated with cutaneous squamous cell carcinoma (cSCC).

Methods: Differentially expressed genes (DEGs) and hub genes in key module were identified using the limma package and weighted gene co-expression network analysis (WGCNA) in R software, respectively. The intersection of these genes was then subjected to LASSO regression to pinpoint characteristic genes. The correlation between immune cell infiltration and these characteristic genes was further elucidated using single-sample Gene Set Enrichment Analysis and Spearman correlation analysis.

Results: A total of 113 DEGs were identified, along with their associated biological pathways. From this pool, five characteristic genes—*ADH1B*, *CCL27*, *ID4*, *LRP4* and *S100A9*—were selected and validated. Immune infiltration analysis revealed significant correlations between these genes and various immune cell types, particularly with *CCL27*, *ID4*, *LRP4* and *S100A9*.

Conclusion: The identification of characteristic genes for cSCC provides valuable insights into its molecular mechanisms. The correlations between these genes and immune cell infiltration suggests their potential roles in tumor immunity.

KEYWORDS

cutaneous squamous cell carcinoma, weighted gene co-expression network analysis, characteristic genes, immune microenvironment (IME), LASSO

Highlights

- Utilizing GO analysis and KEGG enrichment analysis, we delved into the cellular functions and signaling pathways potentially implicated in the differentially expressed genes of cutaneous squamous cell carcinoma.
- We identified five characteristic genes in cutaneous squamous cell carcinoma through integrated analysis of differential expression gene and weighted gene co-expression network analysis.
- Our findings unveiled that four characteristic genes exhibit a strong correlation with infiltrating immune cells, suggesting their potential role in shaping the tumor immune microenvironment.

1 Background

Cutaneous squamous cell carcinoma (cSCC) is the second most common type of non-melanoma skin cancer (NMSC), accounting for approximately 20% of all skin cancer cases (Gong et al., 2018). In the United States, cSCC affects an estimated one million new patients annually, leading to up to 9,000 deaths each year (Karia et al., 2013). Globally, the incidence of cSCC has been rising steadily, with an annual growth rate of 10%–12% (Rogers et al., 2010). This increase poses a significant public health threat, particularly in regions with high ultraviolet exposure, where the incidence is markedly higher. (Green and Olsen, 2017). While most cases of cSCC are curable through surgical excision, approximately 4% of patients experience poor outcomes due to tumor metastasis or local recurrence (Brantsch et al., 2008). The high incidence of cSCC has resulted in a death toll comparable to that of melanoma (Fitzmaurice et al., 2019), imposing substantial physical and economic burdens on patients, a challenge likely to worsen with the aging population.

Early diagnosis and treatment of cSCC remain challenging. Early-stage cSCC often presents as skin nodules or scaly plaques, which can easily be mistaken for other benign conditions, complicating clinical diagnosis (Stratigos et al., 2020a). Although several studies have explored molecular markers for cSCC, there is currently a lack of highly specific and sensitive markers for early diagnosis, limiting their clinical application (Stratigos et al., 2020b). For metastatic cSCC, traditional treatments such as surgery and radiotherapy have shown limited efficacy in advanced stages (Que et al., 2018). Despite recent advances in targeted therapies, effective therapeutic targets for metastatic cSCC remain elusive and treatment outcomes continue to require improvement (Karia et al., 2013; Burton et al., 2016). Therefore, further investigation into the molecular mechanisms underlying cSCC and the identification of reliable biomarkers for early diagnosis and targeted therapies are essential for improving patient outcomes.

Characteristic genes involved in tumorigenesis and progression is crucial for advancing cancer research. These genes, whose abnormal expression or mutations can serve as diagnostic and prognostic biomarkers, also present potential therapeutic targets. For example, the estrogen receptor (*ER*) is a characteristic gene in breast cancer, with approximately 70%–80% of breast cancer patients being ER-positive (ER+), who typically respond well to hormone therapies such as tamoxifen (Davies et al., 2011). In the case of cSCC, mutations in genes such as *TP53*, *NOTCH1* and *CDKN2A* have been implicated in the initiation and progression of the disease (South et al., 2014). However, these mutations are not exclusive to cSCC and are observed in various other cancers. For instance, *TP53* mutations are present in approximately 42% of all cancer patients and are closely associated with tumorigenesis, progression and prognosis (Kandoth et al., 2013). Notably, *TP53* mutations are also detected in 20%–30% of cases of actinic keratosis, a precursor lesion to cSCC (Padilla et al., 2010). This suggests that the currently identified characteristic genes for cSCC lack both sensitivity and specificity. Therefore, further investigation is required to identify more specific and sensitive characteristic genes for cSCC, along with understanding their molecular mechanisms, to improve early diagnosis and targeted therapy.

High-throughput sequencing technologies, combined with weighted gene co-expression network analysis (WGCNA), has

proven effective in identifying characteristic genes in cancer research. High-throughput sequencing enables comprehensive gene expression profiling, while WGCNA identifies disease-related gene modules and hub genes. For instance, in hepatocellular carcinoma, WGCNA and Least Absolute Shrinkage and Selection Operator (LASSO) algorithms have been employed to identify macrophage-related genes and further establish a prognostic model (Wang et al., 2022). While similar approaches have been applied to ovarian cancer to identify fibroblast-related genes and their roles in the tumor microenvironment (Feng et al., 2022). Despite these advances, the application of these techniques to cSCC remains unexplored. In this study, we utilized genome-wide mRNA transcriptomic data from cSCC and normal skin samples available in the Gene Expression Omnibus (GEO) database. By integrating WGCNA and LASSO algorithms, we identified characteristic genes for cSCC. These findings provide a deeper understanding of the molecular mechanisms underlying cSCC.

The tumor immune microenvironment plays a critical role in the initiation and progression of cSCC. Previous studies have demonstrated that tumor-infiltrating lymphocytes in cSCC are closely associated with immune evasion, tumor progression and patient prognosis (Saeidi et al., 2023). Additionally, tumor-associated neutrophils contribute significantly to cSCC growth and immune evasion, thereby facilitating tumor progression (Khou et al., 2020). Investigating the correlation between the expression patterns of characteristic genes and immune cells infiltration provides valuable insights into the molecular mechanisms underlying tumor growth and immune evasion. This approach also offers the potential to identify biomarkers with diagnostic, prognostic, or therapeutic relevance. In the present study, we further explored the relationship between cSCC characteristic genes and immune cell infiltration. Our findings reveal that cSCC characteristic genes are significantly correlated with the infiltration of various immune cell types. These results enhance our understanding of the immune landscape of cSCC.

2 Objects and methods

2.1 Objects

This study utilized data from patients with cutaneous squamous cell carcinoma (cSCC) obtained from the GEO database. Tumor tissue samples were classified as the “Tumor” group, while corresponding normal skin tissue samples were categorized as the “Normal” group for subsequent bioinformatics analysis.

2.2 Methods

2.2.1 Data collection and preprocessing

Gene expression data for cSCC were retrieved from the GEO database using the search terms “Squamous cell carcinoma of skin” AND “*Homo sapiens*”. The inclusion criteria for the datasets were as follows: (1) the dataset contained genome-wide mRNA transcriptomic data, (2) the dataset included samples from both cSCC tissues and normal skin tissues and (3) datasets lacking proper annotation files were excluded. After applying these criteria, five

cSCC microarray datasets were selected: GSE7553, GSE42677, GSE45164, GSE2503 and GSE117247. Among them, GSE7553, GSE42677 and GSE45164 were used as the training set for primary analysis, which included 31 cSCC tissue samples and 17 normal skin tissue samples. GSE2503 and GSE117247 were used as the test set for further validation, which included 13 cSCC tissue samples and 16 normal skin tissue samples. Descriptive statistics for the samples are shown in [Supplementary Table 1](#). Data processing, including matrix organization, imputation, normalization and merging, was performed using the Perl and R programming languages. The limma package from the Bioconductor suite was used for data normalization and integration. Subsequently, the normalized gene expression data were used for analyses.

2.2.2 Identification of differentially expressed genes (DEGs)

DEGs between the Tumor and Normal groups in the training set were identified using the limma package in R. To identify genes with significant expression changes and robust statistical significance, the screening criteria for DEGs were set as $|\log_{2}FC| > 2$ and an adjusted $P < 0.05$, ensuring both biological relevance and statistical significance (Ritchie et al., 2015; Liu et al., 2019). A linear model and eBayes (empirical Bayes adjustment) were applied to enhance statistical robustness. To visualize the DEG results, heatmaps and volcano plots were generated using the pheatmap and ggplot2 R packages. These visualizations were used to highlight the most significantly altered genes between cSCC and normal skin tissue samples.

2.2.3 GO and KEGG pathway enrichment analysis

To further explore the biological functions and underlying mechanisms of the identified DEGs, Gene Ontology (GO) functional enrichment analysis and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis were performed. The analysis was conducted using the clusterProfiler, enrichplot and ggplot2 R packages (Yu et al., 2012). GO analysis provided insights into the biological processes, molecular functions and cellular components associated with the DEGs, while KEGG pathway analysis helped identify the key signaling pathways potentially involved in cSCC development.

2.2.4 WGCNA and identification of hub genes in key module

To explore the gene interactions and their potential associations with cSCC, WGCNA was performed on the normalized gene expression data using the R package “WGCNA”. First, hierarchical clustering (hclust) was employed to construct a sample dendrogram, facilitating the detection and removal of outlier samples. The pickSoftThreshold function was then applied to determine the optimal soft-thresholding power to ensure the network’s scale-free topology. Next, gene modules were identified using the cutreeDynamic function, which employs a dynamic tree-cutting approach. The minimum module size was set to 60, and the cut height was set to 0.25. Modules with high similarity were merged. The relationship between each module and clinical phenotypic data was evaluated by calculating module-trait correlations. A module-trait heatmap was generated to highlight modules significantly associated with cSCC. Hub genes in key modules were selected based on Module Membership (MM) > 0.8 and Gene Significance

(GS) > 0.5 , ensuring that the identified hub genes not only play a crucial role within the module (high MM) but also exhibit a strong correlation with the studied phenotype (high GS) (Langfelder and Horvath, 2008; Ai et al., 2020). The hub genes identified in this manner were then intersected with the DEGs obtained in the previous analysis for further exploration.

2.2.5 LASSO regression analysis

The intersecting genes identified in the previous step were subjected to LASSO regression analysis to further select potential characteristic genes for cSCC. A ten-fold cross-validation approach was used to determine the optimal regularization parameter lambda (λ). The value of λ corresponding to the minimum cross-validation error ($\lambda_{\min} = 0.0003287787$) was selected as the optimal model complexity (Friedman et al., 2010). Genes with non-zero coefficients in the optimized LASSO model were identified as characteristic genes of cSCC.

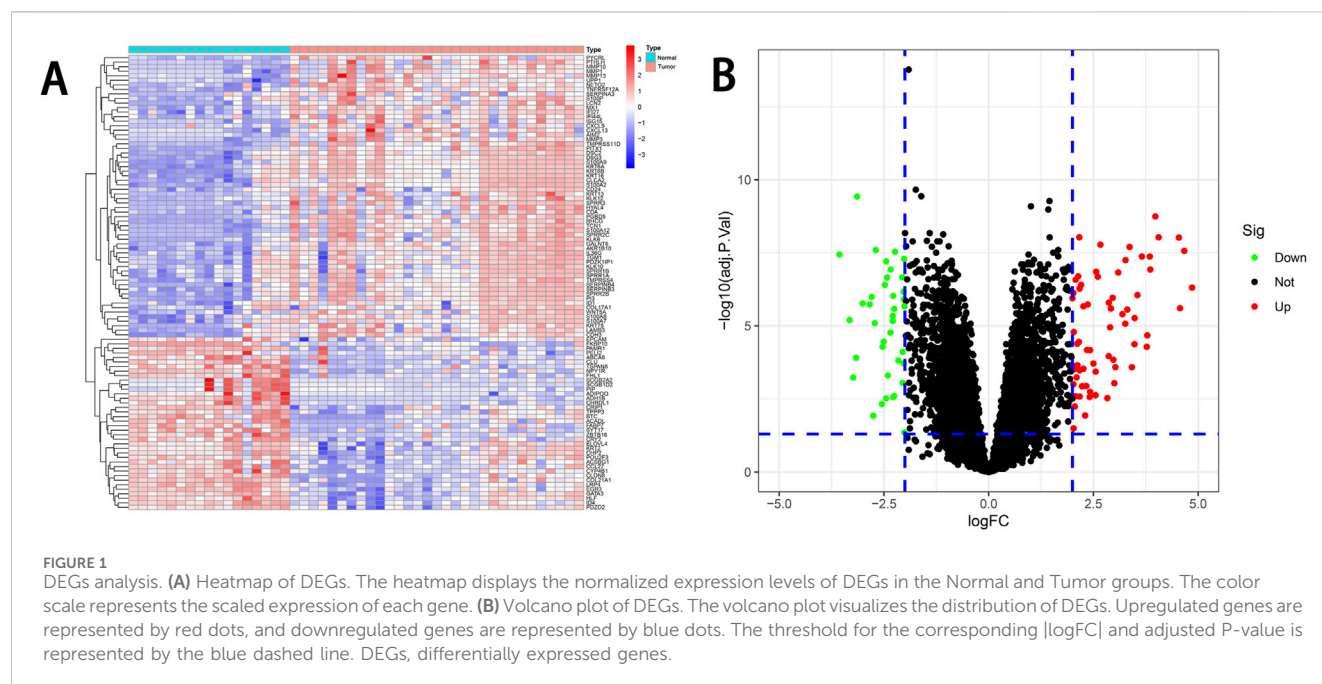
2.2.6 Validation of cSCC characteristic genes

To evaluate the reliability of the identified characteristic genes in cSCC, we conducted comprehensive validation through multiple approaches. Initially, the pROC package in R was used to evaluate the average achieved model performance for ten-fold cross-validation based on the intersecting genes and the characteristic genes in the training set. Subsequently, the overall performance of the model for these characteristic genes in the test set was assessed using the pROC package. Following this, comparative analysis of gene expression profiles between cSCC tumor tissues and normal was performed in both training and test sets using boxplot visualizations created with the ggpvr package.

Additionally, protein-level validation was carried out through immunohistochemical staining of three matched cSCC tumor-adjacent normal tissue pairs, confirming the translational relevance of the identified molecular signatures. Descriptive statistics for the samples are shown in [Supplementary Table 2](#). Tissue sections underwent standard deparaffinization through xylene-ethanol series and microwave-assisted antigen retrieval in citrate buffer (pH 6.0). After cooling to room temperature and phosphate-buffered saline washing (pH 7.4), endogenous peroxidase activity was blocked with 3% H_2O_2 . Non-specific binding was minimized through 30-min incubation with 3% bovine serum albumin (BSA). Sections were incubated overnight at 4°C with primary antibodies against ADH1B (Proteintech), CCL27 (Proteintech), ID4 (Proteintech), LRP4 (Bioss) and S100A9 (Proteintech), followed by room temperature incubation with horseradish peroxidase-conjugated secondary antibodies for 1 h. Color development was achieved using 3,3'-diaminobenzidine with hematoxylin counterstaining. Sections were subsequently dehydrated through ethanol-xylene series and mounted with neutral gum. Staining intensities were quantified as mean optical density using ImageJ software (NIH), normalized to a designated reference sample (set as 1.0 arbitrary unit), and subsequently visualized through GraphPad Prism 10 (GraphPad).

2.2.7 Correlation analysis between cSCC characteristic genes and immune cell infiltration

To explore the association between characteristic genes and immune cell infiltration in cSCC, single-sample Gene Set



Enrichment Analysis (ssGSEA) was used to estimate the relative abundance of 28 immune cell types across all samples. ssGSEA was performed using the normalized expression matrix and an immune gene set (GMT file), generating standardized enrichment scores that reflect the abundance of each immune cell type in each sample. Heatmaps visualizing the ssGSEA scores were constructed using the pheatmap R package. Differences in immune cell infiltration between the Tumor and Normal groups were compared using violin plots, which were created with the vioplot package, and statistical significance was assessed using the Wilcoxon rank-sum test. To further investigate the relationship between characteristic genes and immune cell infiltration, Spearman's rank correlation coefficients (cor) were calculated using the $cor.test$ function in R. The significance of these correlations was determined by P -values, and correlation heatmaps were generated using the ggplot2 package, with annotations to indicate the strength and significance of the associations.

3 Results

3.1 Identification of DEGs

We conducted a comprehensive analysis of the preprocessed gene expression data using R, resulting in the identification of 113 DEGs. Among these, 38 genes were downregulated, and 75 genes were upregulated. The differential expression of these genes was visualized using heatmaps and volcano plots (Figure 1).

3.2 Functional enrichment analysis of DEGs

To further explore the biological roles and pathways of the identified DEGs, we performed GO and KEGG enrichment analyses using the ClusterProfiler package. The DEGs were categorized and

analyzed in terms of biological processes, cellular components and molecular functions. In terms of biological processes, the DEGs were predominantly associated with keratinization, keratinocyte differentiation, epidermal development, epithelial cell differentiation and skin development (Figures 2A–C). These findings indicate that these genes play critical roles in skin development, keratinization, and cell differentiation. With respect to cellular components, the DEGs were enriched in the cornified envelope, secretory granule lumen, cytoplasmic vesicle lumen, vesicle lumen and keratin filaments (Figures 2A–C), underscoring their significant roles in keratinization and secretion-related processes. Regarding molecular functions, the DEGs were involved in activities such as structural constituent of skin epidermis, RAGE receptor binding, calcium-dependent protein binding, serine-type endopeptidase activity and serine-type peptidase activity (Figures 2A–C).

Furthermore, KEGG pathway analysis revealed that the DEGs were predominantly implicated in several pathways, including the IL-17 signaling pathway, PPAR signaling pathway, viral protein interaction with cytokines and cytokine receptors and amoebiasis (Figures 2D–F).

3.3 Identification of hub genes using WGCNA

To identify biologically significant characteristic genes, uncover potential regulatory mechanisms of cSCC, narrow the research focus, and enhance analytical efficiency, we applied WGCNA to identify gene modules and hub genes that were highly correlated with specific phenotypes.

Initially, hierarchical clustering was performed on the gene expression data to construct a sample dendrogram. No outlier samples were detected, indicating that the samples exhibited a concentrated clustering pattern and the data quality was high,

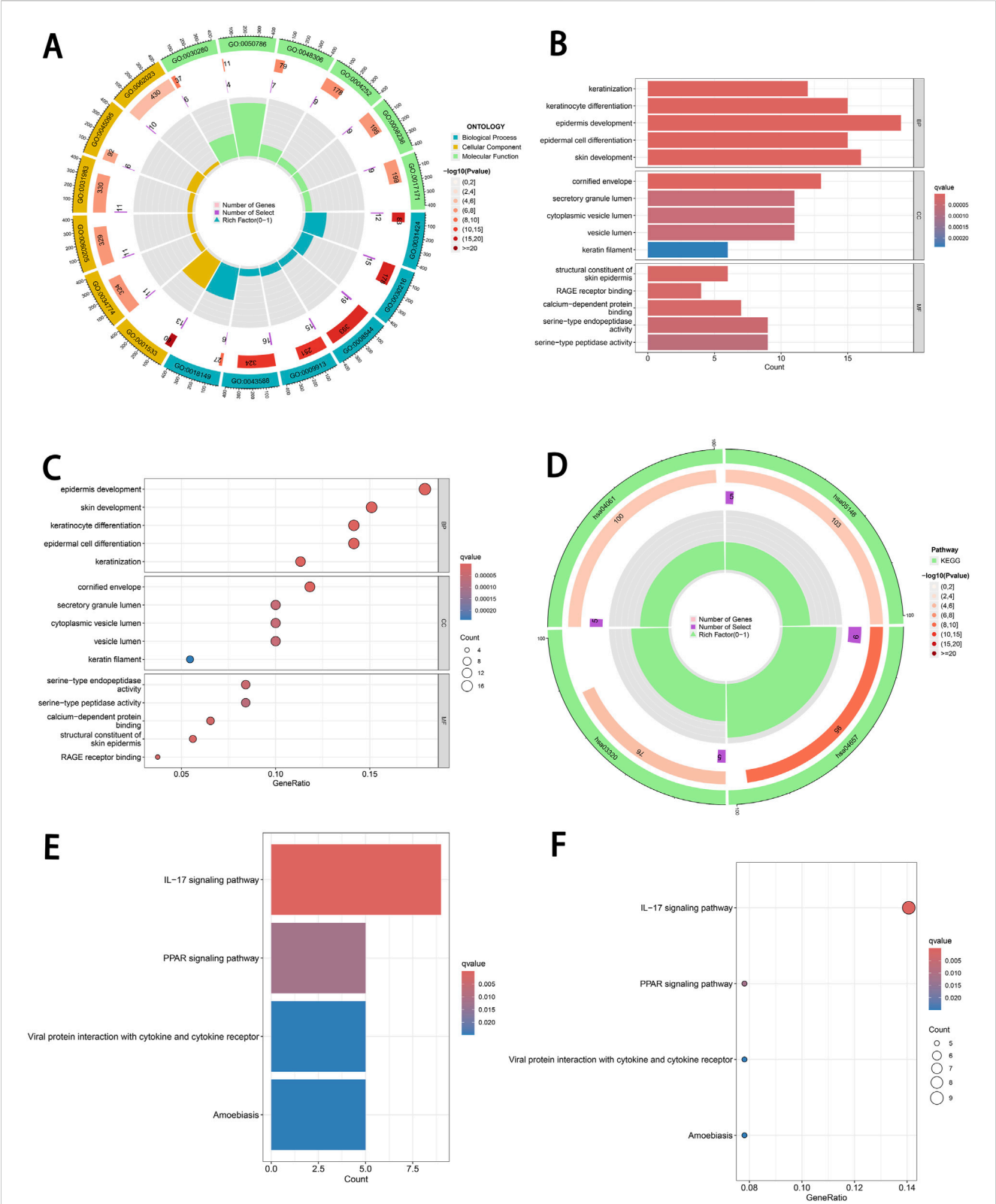


FIGURE 2 Functional enrichment analysis of DEGs. **(A)** Circular plot of GO enrichment. This plot illustrates the enriched GO terms associated with DEGs in three categories: BP, CC and MF. **(B)** Bar plot of GO enrichment. GO terms are ranked by significance ($-\log_{10}(P\text{-value})$). **(C)** Bubble plot of GO enrichment. The size of each bubble represents the number of DEGs associated with a specific GO term, while the color gradient reflects the q-value. **(D)** Circular plot of KEGG pathway enrichment. This plot displays the KEGG pathways significantly enriched by DEGs. **(E)** Bar plot of KEGG pathway enrichment. Pathways are ranked by $-\log_{10}(P\text{-value})$, with bars showing the corresponding q-values and the number of DEGs associated with each pathway. **(F)** Bubble plot of KEGG pathway enrichment. Bubble size indicates the number of DEGs associated with each pathway, while the color gradient reflects q-value significance. DEGs, differentially expressed genes. GO, Gene Ontology. BP, biological processes. CC, cellular components. MF, molecular functions.

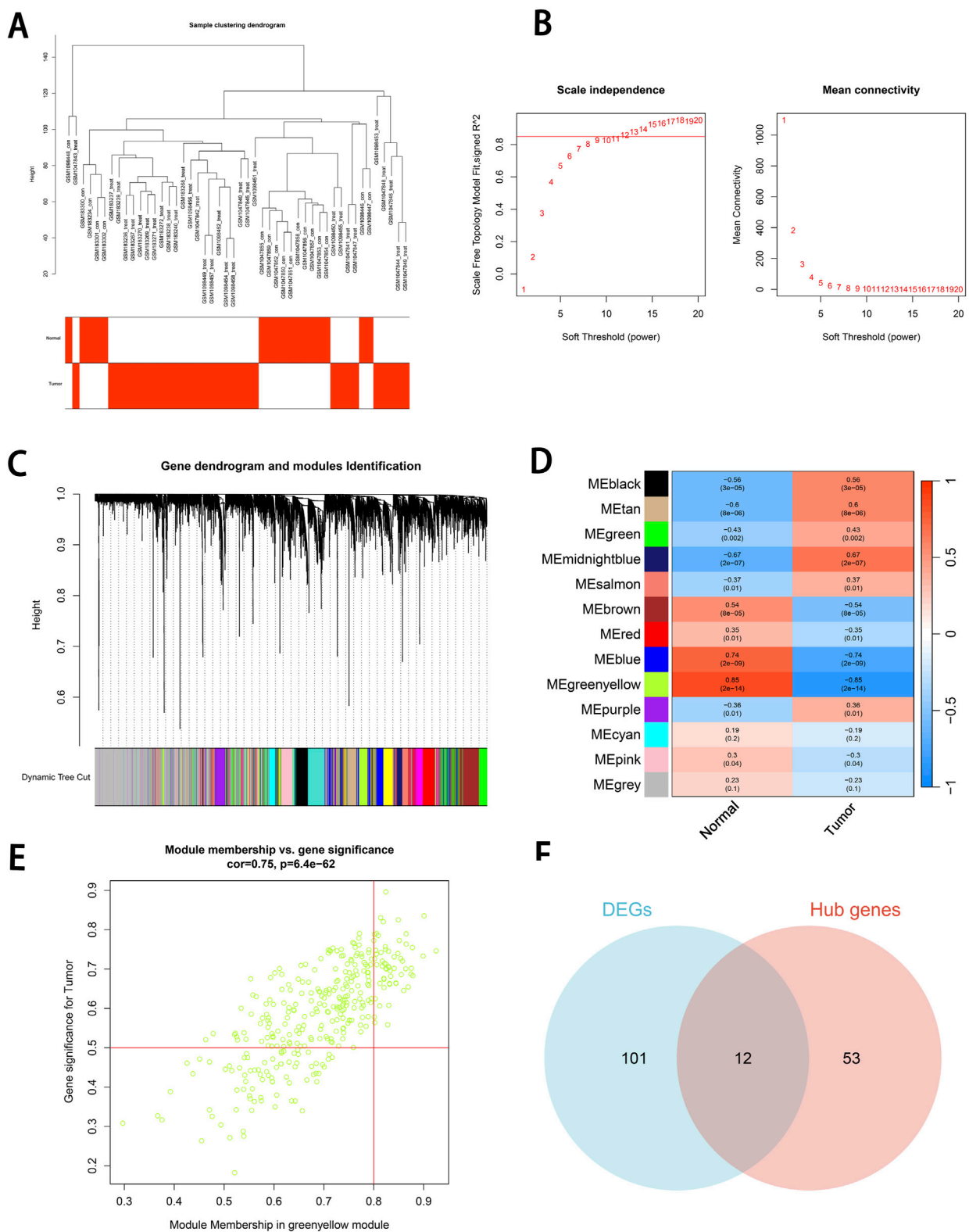


FIGURE 3 Co-expression module construction and identification of key module and hub genes using WGCNA. **(A)** Sample clustering dendrogram. Hierarchical clustering is performed based on gene expression data. **(B)** Scale Independence and mean connectivity. The scale-free topology fit index (R^2) corresponding to soft thresholding powers (β) from 1 to 20 is shown in the left panel, and the average connectivity (k) of β values from 1 to 20 is shown in the right panel. **(C)** Gene dendrogram and modules Identification. Differentially expressed genes are clustered based on the dynamic tree cutting method, with representing different gene modules are represented by different colors. **(D)** Identification of key module. The correlation between 13 modules and two traits (Normal group and Tumor group) is shown in the figure. The correlation coefficient and significance level are indicated by the numerical values. The scaled correlations are represented by the color scale. **(E)** Identification of hub genes. A scatter plot between the module *(Continued)*

FIGURE 3 (Continued)

membership (MM) in the green-yellow module and the gene significance (GS) for cSCC is displayed in the figure. Genes in the upper right corner are selected as hub genes. (F) Venn diagram of intersecting genes. The intersecting genes between DEGs and hub genes are shown in the Venn diagram. GS, gene significance. MM, module membership.

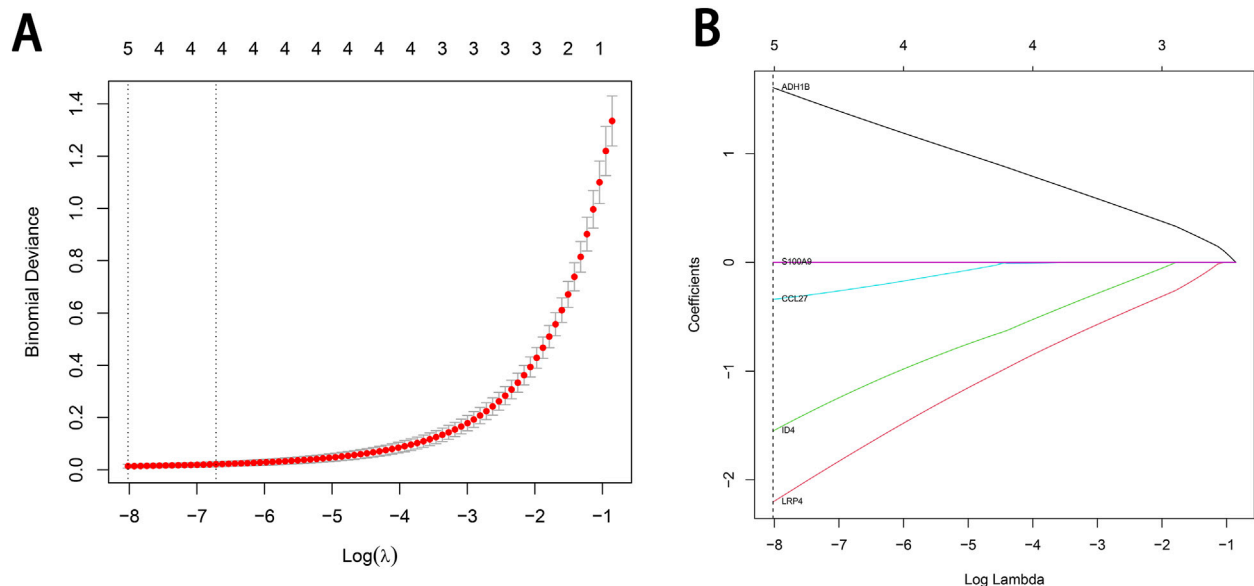


FIGURE 4

LASSO regression analysis for the selection of characteristic genes in cSCC. (A) The binomial deviance plot at different $\log(\lambda)$ values. The relationship between $\log(\lambda)$ and the model fit is shown in the figure. The optimal λ value determined by cross-validation is indicated by the dashed line on the left. (B) The gene coefficient curve plot at different $\log(\lambda)$ values. The relationship between the coefficient values of different genes and $\log(\lambda)$ is shown in the figure, and the figure has been marked with the selected feature genes at the optimal λ value.

providing a reliable foundation for subsequent analysis (Figure 3A). The optimal soft-thresholding power was determined to be $\beta = 12$, ensuring that the network met the scale-free topology criterion ($R^2 = 0.85$) while maintaining high gene connectivity. The high scale-free topology value indicated that the gene expression network adhered to the characteristics of biological networks (Figure 3B).

Subsequently, the minimum module size was set to 60, and the cut height was set to 0.25. Using the dynamic tree-cutting method, the genes were grouped into 13 modules, with each module color reflecting the similarity in gene expression patterns. These modules served as the foundation for exploring phenotype-related modules (Figure 3C). The correlation between genes within each module and the phenotypic data was calculated, and a module-trait heatmap was generated. This analysis revealed that the green-yellow module exhibited a strong positive correlation with cSCC ($r = 0.85$, $P < 0.001$), suggesting its potential as a key module for further analysis (Figure 3D).

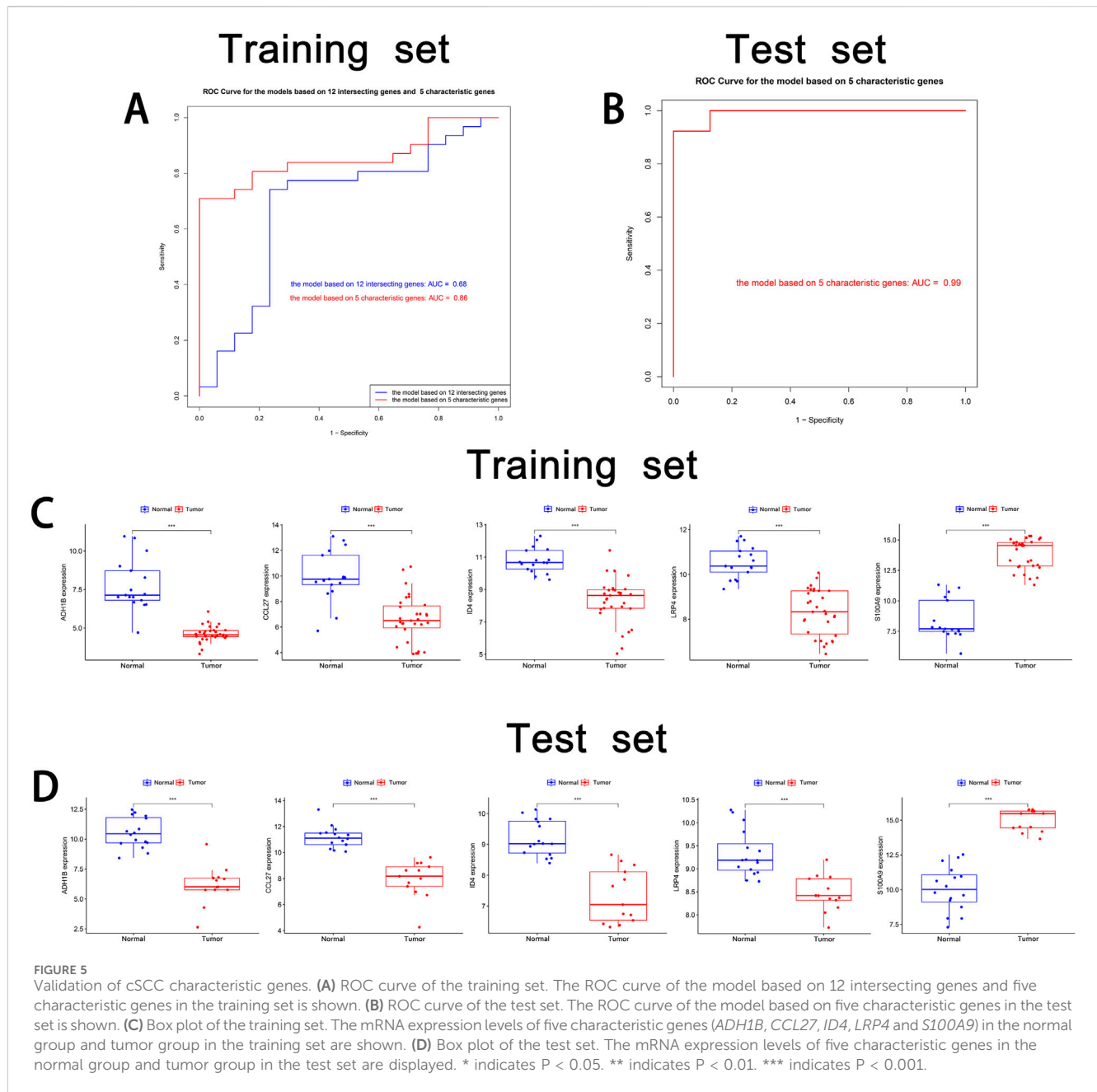
Hub genes within the green-yellow module were identified based on thresholds for $MM > 0.8$ and $GS > 0.5$. We identified 65 hub genes, which not only played a crucial role within the module but also exhibited strong correlations with the phenotype of interest (Figure 3E). The hub genes were then intersected with the DEGs, and a Venn diagram was generated. This analysis revealed 12 intersecting genes that were both highly correlated within the green-yellow module and differentially expressed (Figure 3F).

3.4 LASSO regression analysis

To identify characteristic genes associated with cSCC while minimizing the risk of overfitting, we performed LASSO regression combined with ten-fold cross-validation on the expression levels of the 12 intersecting genes. Genes with non-zero regression coefficients were selected, and the results were visualized (Figures 4A, B). The 12 intersecting genes and their coefficients, which reflect the significance of each gene, obtained from the LASSO regression are presented in Supplementary Table 3. Ultimately, five genes—Alcohol Dehydrogenase 1B (*ADH1B*), C-C Motif Chemokine Ligand 27 (*CCL27*), Inhibitor of DNA Binding 4 (*ID4*), LDL Receptor Related Protein 4 (*LRP4*) and S100 Calcium Binding Protein A9 (*S100A9*)—were identified as characteristic genes of cSCC.

3.5 Validation of cSCC characteristic genes

To validate the characteristic genes, the receiver operating characteristic (ROC) curves on the training set were used to assess the overall performance of the models based on 12 intersecting genes and five characteristic genes. The model based on 12 intersecting genes had an Area Under the Curve (AUC) value of 0.68, while the model based on five characteristic genes achieved a higher AUC of 0.86 (Figure 5A). This



demonstrated that the model after combined LASSO regression offers superior predictive performance. Additional performance metrics, such as sensitivity and specificity are provided in [Supplementary Table 4](#). Subsequent external validation on the test set revealed exceptional generalizability, demonstrating robust discriminative capacity with an AUC of 0.99 ([Figure 5B](#)).

mRNA expression analysis, visualized using boxplots generated by ggpubr, revealed consistent expression patterns between the training and test sets for all characteristic genes, with statistically significant differential expression ([Figures 5C, D](#)). This finding confirmed the stable identification of these genes across distinct patient cohorts. Notably, to corroborate the transcriptional findings at protein level, we conducted comparative immunohistochemical analysis between cSCC lesions and matched normal skin tissues. The protein expression profiles demonstrated distinct differential patterns

that mirrored the mRNA expression trends ([Figure 6](#)). Quantitative evaluation revealed consistent downregulation of *ADH1B*, *CCL27*, *ID4* and *LRP4* in tumor tissues compared to normal counterparts. Conversely, *S100A9* exhibited marked upregulation in tumor. This multi-level concordance between transcriptional and translational data strongly corroborates the biological relevance of these characteristic genes in cSCC development.

3.6 Correlation analysis between cSCC characteristic genes and infiltrating immune cells

Although the five characteristic genes were identified, their exact roles in cSCC tumorigenesis and progression remain unclear. Given

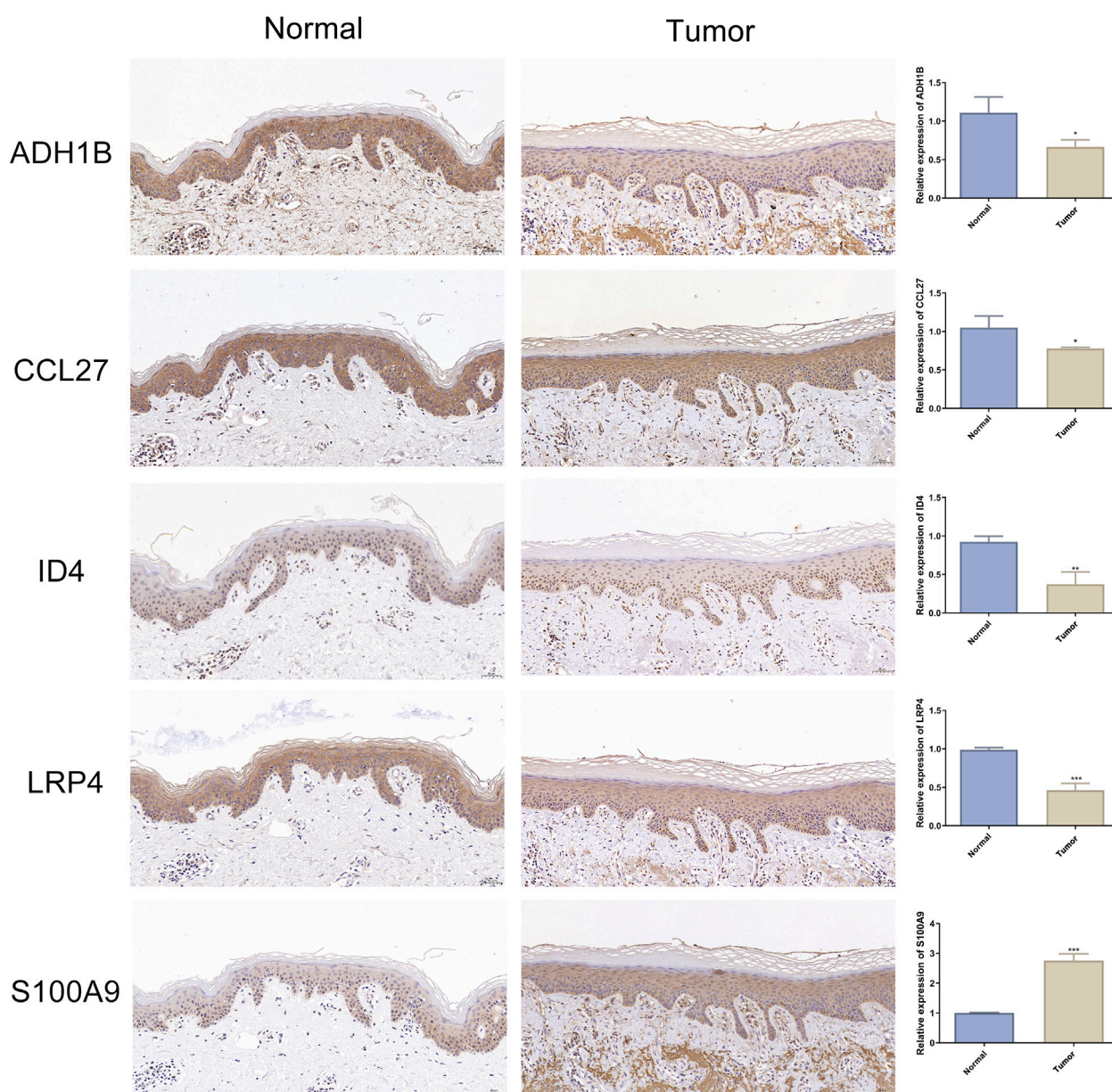


FIGURE 6

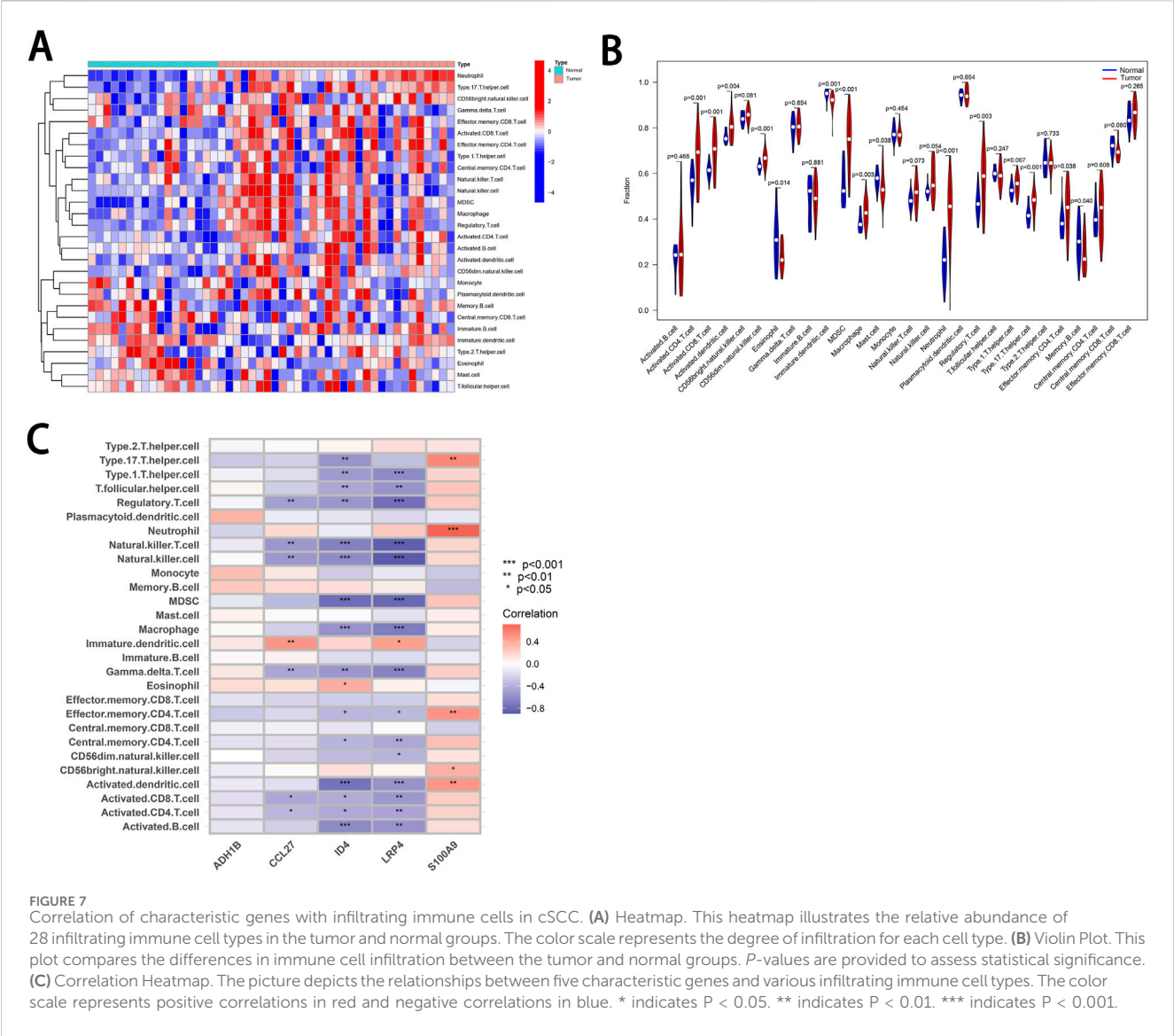
Expression of cSCC characteristic genes in immunohistochemistry. Representative immunohistochemical images of ADH1B, CCL27, ID4, LRP4 and S100A9 expression in adjacent normal skin (left panel) and cSCC tumor (middle panel). Nuclei were counterstained with hematoxylin (blue), with positive signals shown by 3,3'-diaminobenzidine deposition (brown). Scale bars: 50 μ m. Quantitative analysis of normalized expression levels (right panel). Data expressed as Mean \pm SD. $n = 3$. * indicates $P < 0.05$. ** indicates $P < 0.01$. *** indicates $P < 0.001$.

the established connection between the immune microenvironment and cSCC development (Saeidi et al., 2023), and the functional enrichment analysis indicating the involvement of these genes in immune-related pathways (e.g., *S100A9* in the IL-17 signaling pathway as shown in Supplementary Table 5), we further explored the relationship between these genes and the infiltration of immune cells.

Using the ssGSEA method, we analyzed the infiltration of 28 immune cell types in the Tumor and Normal groups. The results revealed statistically significant differences in the infiltration of 14 immune cell types (Supplementary Table 6). Among these, 10 immune cell types exhibited higher levels of

infiltration in the Tumor group, while five showed significantly lower levels in the Tumor group compared to the Normal group (Figures 7A, B).

Furthermore, we performed a correlation analysis between the cSCC characteristic genes and the abundance of infiltrating immune cells (Figure 7C). The analysis revealed that four of the characteristic genes—*CCL27*, *ID4*, *LRP4* and *S100A9*—were significantly associated with various types of infiltrating immune cells. For instance, *S100A9* expression was positively correlated with the abundance of Type 17 T helper cells, neutrophils, effector memory CD4 T-cell, CD56 bright natural killer cells and activated dendritic cells. The correlations between the five



characteristic genes and various immune cell types, along with their corresponding *P*-values, are provided in [Supplementary Table 7](#). These results suggest that the characteristic genes *CCL27*, *ID4*, *LRP4* and *S100A9* may play a role in modulating the immune microenvironment of cSCC.

4 Discussion

cSCC, a major subtype of NMSC, is characterized by its potential for local invasion and distant metastasis. Although the overall metastatic rate is relatively low, with approximately 4% of patients developing lymph node metastasis and 1.5% succumbing to the disease (Karia et al., 2013), advanced cSCC remains a significant clinical challenge. Early-stage cSCC can be effectively treated with current therapies, but management of advanced cases continues to pose considerable difficulties. Understanding the characteristic genes involved in cSCC pathogenesis is critical for elucidating the mechanisms underlying its development and

progression, and for advancing diagnostic and therapeutic strategies. However, research into the molecular profiles of cSCC remains limited, and systematic investigations to identify and explore characteristic genes are urgently needed.

In the present study, we identified five characteristic genes from three cSCC datasets (GSE45164, GSE42677 and GSE7553) for the first time. While the original study of GSE42677 also identified *S100A9* as a DEG, our combined approach, which includes WGCNA, LASSO regression analysis and immunohistochemical staining, further validates that *S100A9*, along with the other four genes, collectively serves as a set of characteristic genes for cSCC. Due to significant differences in research objectives, analytical methods, sample sizes and sample compositions, the conclusions drawn in our study exhibit notable disparities from those of the original studies.

Among these characteristic genes, *ADH1B*, *CCL27*, *ID4* and *LRP4* were found to be significantly downregulated in cSCC (Figure 5C). Previous studies have highlighted the roles of these genes in the tumorigenesis and progression of other cancers,

suggesting that their altered expression may be crucial for the development of cSCC. For instance, polymorphisms in the *ADH1B* gene, such as rs1229984, have been shown to correlate with a reduced risk of head and neck cancer (Imani et al., 2024). Additionally, *ADH1B* expression is notably reduced in hepatocellular carcinoma (HCC), with low expression levels associated with poor prognosis (Liu et al., 2020). Although our study does not provide direct evidence linking *ADH1B* to cSCC prognosis, its downregulation suggests that it may be involved in disease progression, warranting further investigation in clinical cohorts. *CCL27*, a chemokine involved in immune cell recruitment and lymphangiogenesis, has been shown to be upregulated in several cancers, including breast cancer and melanoma, where it contributes to tumor metastasis by promoting lymphatic spread (Karnezis et al., 2019). However, this study indicates that *CCL27* is downregulated in cSCC. Given that cSCC exhibits significantly lower lymph node metastatic potential compared to breast cancer and melanoma, this finding suggests that *CCL27* may play a role in the differential lymphatic metastasis capabilities among various cancers. Further research is needed to elucidate the molecular and biological functions of *CCL27* in tumor metastasis. The expression of *ID4* has been reported to be suppressed by UVB irradiation via DNA methylation, which in turn promotes cell proliferation, migration and invasion, contributing to tumor formation in mouse models of skin cancer (Li et al., 2020). In non-small cell lung cancer (NSCLC), *ID4* upregulation is associated with improved prognosis, suggesting its potential role as a tumor suppressor (Rodón et al., 2019). Similarly, *ID4* was found to be downregulated in cSCC in our study. It remains unclear whether this downregulation contributes to cSCC progression, and further experimental validation is needed to explore its potential role. *LRP4*, a member of the low-density lipoprotein receptor family, has been implicated in various cancers. In NSCLC, *LRP4* downregulation promotes tumorigenesis and progression by modulating the Wnt/ β -catenin signaling pathway (Zhang et al., 2024). In contrast, *LRP4* is upregulated in papillary thyroid carcinoma, where it promotes cell proliferation, migration and invasion (Zhou et al., 2018). This contradictory expression pattern suggests that *LRP4* may have context-dependent roles in different cancer types, potentially influenced by the tumor microenvironment. However, the prognostic significance of *LRP4* in cSCC remains to be determined, and further studies are needed to explore its functional role. In summary, our findings reveal that four of the identified characteristic genes (*ADH1B*, *CCL27*, *ID4* and *LRP4*) are consistently downregulated in cSCC, and prior studies suggest their involvement in tumor proliferation, migration and invasion. However, the specific roles of these genes in cSCC require further clinical and experimental validation.

In addition to the four downregulated characteristic genes mentioned, this study also confirmed the upregulation of *S100A9* in cSCC (Figure 5C). *S100A9* is a member of the S100 protein family, which binds calcium and plays a critical role in the recruitment and adhesion of immune cells such as neutrophils and monocytes, especially in inflammatory responses and inflammation-associated diseases (Wang et al., 2018). Previous studies have demonstrated that *S100A9* is upregulated in various cancers and is closely associated with tumor initiation, progression and prognosis (Acharyya et al., 2012; Eisenblaetter et al., 2017). In

cSCC, elevated *S100A9* expression has been found to bind to Receptor for Advanced Glycation Endproducts on myeloid-derived suppressor cells (MDSCs), activating NF- κ B signaling. This interaction promotes the recruitment of MDSCs to the tumor site, which suppresses the immune system's anti-tumor response, thereby facilitating tumor growth and metastasis (Markowitz and Carson, 2013). Our study also observed a potential positive correlation between *S100A9* and MDSC infiltration (Figure 7C). While the p-value was not sufficient to confirm statistical significance, the results partially support previous findings. Interestingly, downregulation of *S100A9* expression in cSCC has been shown to inhibit tumor proliferation and migration (Zhang et al., 2021). Furthermore, IL-6 has been clinically associated with poor prognosis in cSCC (Mallardo et al., 2023), thereby underscoring the critical involvement of inflammatory molecules in tumor progression. Future investigations into how these inflammatory mediators collaboratively orchestrate tumor development through synergistic molecular networks will be of paramount importance for advancing therapeutic strategies.

Notably, the intricate tumor immune microenvironment involves not only MDSC-mediated immunosuppression but also dynamic equilibrium among immune cell populations. Our study revealed concurrent infiltration of functionally antagonistic CD8⁺ cytotoxic T-cell and immunosuppressive Tregs within tumor tissues, a paradoxical coexistence that has been previously documented in renal clear cell carcinoma and lung adenocarcinoma (Pan et al., 2020; Meng et al., 2024). Current research demonstrates functional heterogeneity in tumor-infiltrating CD8⁺ T-cell, where many exhibit non-reactive "bystander" phenotypes that challenge the prognostic utility of quantitative assessments alone (Simoni et al., 2018). Moreover, Treg cell-mediated immunosuppression further impairs the functional state of CD8⁺ T-cell (Sakaguchi et al., 2008). These findings collectively suggest that effective immunotherapeutic strategies should adopt dual targeting approaches: enhancing effector T-cell functionality while selectively inhibiting Treg-mediated immunosuppression to achieve durable antitumor responses.

Furthermore, in this study, we established correlations between the cSCC characteristic genes and various infiltrating immune cell populations through bioinformatics analysis (Figure 7C). These findings suggest that these genes may play a role in shaping the cSCC immune microenvironment and promoting immune evasion, a hypothesis supported by previous studies. For instance, *S100A9* is highly expressed in neutrophils and monocytes and acts as a promoter by regulating tumor metabolism and the immune microenvironment, thereby facilitating tumor growth and metastasis (Chen et al., 2023). Further research has revealed that tumor-associated neutrophils can suppress CD8⁺ T-cell function and express Programmed Death-Ligand 1, contributing to immune evasion and promoting cSCC progression (Khou et al., 2020). Our study reaffirms the positive correlation between high *S100A9* expression and increased neutrophil infiltration in cSCC, consistent with previous reports. While there is substantial evidence supporting the involvement of *S100A9* in the tumor immune microenvironment and immune evasion, the roles of *CCL27*, *ID4* and *LRP4*, genes that also showed strong correlations with infiltrating immune cells, remain inadequately explored. Future

research is needed to clarify whether these genes influence the tumor immune environment and to elucidate their specific mechanisms. Such studies will deepen our understanding of cSCC and could uncover novel targets for immunotherapy.

However, this study has several limitations. First, the available datasets that met the analysis criteria were limited, resulting in a relatively small sample size that may not fully capture the heterogeneity of cSCC. Second, although multiple screening methods were employed, the stability and generalizability of the identified characteristic genes require validation in larger, multi-center cohorts. Additionally, the prognostic significance of these genes in cSCC has not been systematically evaluated, which should be addressed in future studies. Future research should focus on expanding sample sizes, incorporating diverse data sources, conducting functional validation experiments, integrating multi-omics data, and performing prospective clinical studies to improve the representativeness and clinical applicability of the findings.

In conclusion, this study employed bioinformatics techniques, including WGCNA, to identify five characteristic genes associated with cSCC: *ADH1B*, *CCL27*, *ID4*, *LRP4* and *S100A9*. Additionally, we demonstrated that *CCL27*, *ID4*, *LRP4* and *S100A9* are correlated with the infiltration of various immune cell types in cSCC. Further investigation into the roles and mechanisms of these characteristic genes in cSCC formation, progression, migration and immune evasion will enhance our understanding of the disease and may provide a foundation for future diagnostic biomarkers and therapeutic targets.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: <https://www.ncbi.nlm.nih.gov/geo/>, GSE42677; <https://www.ncbi.nlm.nih.gov/geo/>, GSE45164; <https://www.ncbi.nlm.nih.gov/geo/>, GSE7553.

Author contributions

QX: Data curation, Investigation, Methodology, Software, Writing – original draft. PX: Data curation, Software, Visualization, Writing – review and editing. WX: Data curation, Investigation, Methodology, Software, Writing – review and editing. QS: Funding acquisition, Investigation, Methodology, Resources,

Supervision, Writing – review and editing. YM: Conceptualization, Data curation, Investigation, Methodology, Writing – review and editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. The study was funded by the 2023 Provincial Health Commission Science and Technology Plan Project (202311522).

Acknowledgments

We thank the grants from the Health Commission Science and Technology Plan Foundation of Jiangxi Province, China. We thank the GEO database for its platforms. Special thanks to Chaowen Zhang, Wuhua Tong and Ting Li for their invaluable assistance and technical support in the preparation and staining of pathological slides.

Conflict of interest

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

Publisher's note

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

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2025.1470584/full#supplementary-material>

References

- Acharyya, S., Oskarsson, T., Vanharanta, S., Malladi, S., Kim, J., Morris, P. G., et al. (2012). A CXCL1 paracrine network links cancer chemoresistance and metastasis. *Cell* 150, 165–178. doi:10.1016/j.cell.2012.04.042
- Ai, D., Wang, Y., Li, X., and Pan, H. (2020). Colorectal cancer prediction based on weighted gene Co-expression network analysis and variational auto-encoder. *Biomolecules* 10, 1207. doi:10.3390/biom10091207
- Brantsch, K. D., Meisner, C., Schönfisch, B., Trilling, B., Wehner-Caroli, J., Röcken, M., et al. (2008). Analysis of risk factors determining prognosis of cutaneous squamous-cell carcinoma: a prospective study. *Lancet Oncol.* 9, 713–720. doi:10.1016/S1470-2045(08)70178-5
- Burton, K. A., Ashack, K. A., and Khachemoune, A. (2016). Cutaneous squamous cell carcinoma: a review of high-risk and metastatic disease. *Am. J. Clin. Dermatol.* 17, 491–508. doi:10.1007/s40257-016-0207-3
- Chen, Y., Ouyang, Y., Li, Z., Wang, X., and Ma, J. (2023). S100A8 and S100A9 in cancer. *Biochim. Biophys. Acta Rev. Cancer* 1878, 188891. doi:10.1016/j.bbcan.2023.188891
- Davies, C., Godwin, J., Gray, R., Clarke, M., Cutter, D., Darby, S., et al. (2011). Relevance of breast cancer hormone receptors and other factors to the efficacy of adjuvant tamoxifen: patient-level meta-analysis of randomised trials. *Lancet* 378, 771–784. doi:10.1016/S0140-6736(11)60993-8
- Eisenblaetter, M., Flores-Borja, F., Lee, J. J., Wefers, C., Smith, H., Hueting, R., et al. (2017). Visualization of tumor-immune interaction - target-specific imaging of S100A8/A9 reveals pre-metastatic niche establishment. *Theranostics* 7, 2392–2401. doi:10.7150/thno.17138
- Feng, S., Xu, Y., Dai, Z., Yin, H., Zhang, K., and Shen, Y. (2022). Integrative analysis from multicenter studies identifies a WGCNA-derived cancer-associated fibroblast

signature for ovarian cancer. *Front. Immunol.* 13, 951582. doi:10.3389/fimmu.2022.951582

Fitzmaurice, C., Abate, D., Abbasi, N., Abbastabar, H., Abd-Allah, F., Abdel-Rahman, O., et al. (2019). Global, regional, and national cancer incidence, mortality, years of life lost, years lived with disability, and disability-adjusted life-years for 29 cancer groups, 1990 to 2017: a systematic analysis for the global burden of disease study. *JAMA Oncol.* 5, 1749–1768. doi:10.1001/jamaoncol.2019.2996

Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* 33, 1–22. doi:10.18637/jss.v033.i01

Gong, T., Yu, Y., Yang, B., Lin, M., Huang, J.-W., Cheng, B., et al. (2018). Celecoxib suppresses cutaneous squamous-cell carcinoma cell migration via inhibition of SDF1-induced endocytosis of CXCR4. *Onco Targets Ther.* 11, 8063–8071. doi:10.2147/OTT.S180472

Green, A. C., and Olsen, C. M. (2017). Cutaneous squamous cell carcinoma: an epidemiological review. *Br. J. Dermatol.* 177, 373–381. doi:10.1111/bjd.15324

Imani, M. M., Moradi, M. M., Rezaei, F., Mozaffari, H. R., Sharifi, R., Safaei, M., et al. (2024). Association between alcohol dehydrogenase polymorphisms (rs1229984, rs1573496, rs1154460, and rs284787) and susceptibility to head and neck cancers: a systematic review and meta-analysis. *Arch. Oral Biol.* 160, 105898. doi:10.1016/j.archoralbio.2024.105898

Kandoth, C., McLellan, M. D., Vandin, F., Ye, K., Niu, B., Lu, C., et al. (2013). Mutational landscape and significance across 12 major cancer types. *Nature* 502, 333–339. doi:10.1038/nature12634

Karia, P. S., Han, J., and Schmuls, C. D. (2013). Cutaneous squamous cell carcinoma: estimated incidence of disease, nodal metastasis, and deaths from disease in the United States, 2012. *J. Am. Acad. Dermatol.* 68, 957–966. doi:10.1016/j.jaad.2012.11.037

Karneizis, T., Farnsworth, R. H., Harris, N. C., Williams, S. P., Caesar, C., Byrne, D. J., et al. (2019). CCL27/CCL28-CCR10 chemokine signaling mediates migration of lymphatic endothelial cells. *Cancer Res.* 79, 1558–1572. doi:10.1158/0008-5472.CAN-18-1858

Khou, S., Popa, A., Luci, C., Bihl, F., Meghraoui-Kheddar, A., Bourdely, P., et al. (2020). Tumor-associated neutrophils dampen adaptive immunity and promote cutaneous squamous cell carcinoma development. *Cancers (Basel)* 12, 1860. doi:10.3390/cancers12071860

Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. *BMC Bioinforma.* 9, 559. doi:10.1186/1471-2105-9-559

Li, L., Li, F., Xia, Y., Yang, X., Lv, Q., Fang, F., et al. (2020). UVB induces cutaneous squamous cell carcinoma progression by *de novo* ID4 methylation via methylation regulating enzymes. *EBioMedicine* 57, 102835. doi:10.1016/j.ebiom.2020.102835

Liu, H., Chen, D., Liu, P., Xu, S., Lin, X., and Zeng, R. (2019). Secondary analysis of existing microarray data reveals potential gene drivers of cutaneous squamous cell carcinoma. *J. Cell Physiol.* 234, 15270–15278. doi:10.1002/jcp.28172

Liu, X., Li, T., Kong, D., You, H., Kong, F., and Tang, R. (2020). Prognostic implications of alcohol dehydrogenases in hepatocellular carcinoma. *BMC Cancer* 20, 1204. doi:10.1186/s12885-020-07689-1

Mallardo, D., Simeone, E., Festino, L., Tuffanelli, M., Vanella, V., Trojaniello, C., et al. (2023). IL-6 as new prognostic factor in patients with advanced cutaneous squamous cell carcinoma treated with cemiplimab. *J. Transl. Med.* 21, 140. doi:10.1186/s12967-023-03971-5

Markowitz, J., and Carson, W. E., 3rd (2013). Review of S100A9 biology and its role in cancer. *Biochim. Biophys. Acta* 1835, 100–109. doi:10.1016/j.bbcan.2012.10.003

Meng, X., Zhao, X., Zhou, B., Song, W., Liang, Y., Liang, M., et al. (2024). FSTL3 is associated with prognosis and immune cell infiltration in lung adenocarcinoma. *J. Cancer Res. Clin. Oncol.* 150, 17. doi:10.1007/s00432-023-05553-w

Padilla, R. S., Sebastian, S., Jiang, Z., Nindl, I., and Larson, R. (2010). Gene expression patterns of normal human skin, actinic keratosis, and squamous cell carcinoma: a spectrum of disease progression. *Arch. Dermatol.* 146, 288–293. doi:10.1001/archdermatol.2009.378

Pan, Q., Wang, L., Chai, S., Zhang, H., and Li, B. (2020). The immune infiltration in clear cell Renal Cell Carcinoma and their clinical implications: a study based on TCGA and GEO databases. *J. Cancer* 11, 3207–3215. doi:10.7150/jca.37285

Que, S. K. T., Zwald, F. O., and Schmuls, C. D. (2018). Cutaneous squamous cell carcinoma: incidence, risk factors, diagnosis, and staging. *J. Am. Acad. Dermatol.* 78, 237–247. doi:10.1016/j.jaad.2017.08.059

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., et al. (2015). Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 43, e47. doi:10.1093/nar/gkv007

Rodón, L., Svensson, R. U., Wiater, E., Chun, M. G. H., Tsai, W.-W., Eichner, L. J., et al. (2019). The CREB coactivator CRTC2 promotes oncogenesis in LKB1-mutant non-small cell lung cancer. *Sci. Adv.* 5, eaaw6455. doi:10.1126/sciadv.aaw6455

Rogers, H. W., Weinstock, M. A., Harris, A. R., Hinkley, M. R., Feldman, S. R., Fleischer, A. B., et al. (2010). Incidence estimate of nonmelanoma skin cancer in the United States, 2006. *Arch. Dermatol.* 146, 283–287. doi:10.1001/archdermatol.2010.19

Saeidi, V., Doudican, N., and Carucci, J. A. (2023). Understanding the squamous cell carcinoma immune microenvironment. *Front. Immunol.* 14, 1084873. doi:10.3389/fimmu.2023.1084873

Sakaguchi, S., Yamaguchi, T., Nomura, T., and Ono, M. (2008). Regulatory T cells and immune tolerance. *Cell* 133, 775–787. doi:10.1016/j.cell.2008.05.009

Simoni, Y., Becht, E., Fehlings, M., Loh, C. Y., Koo, S.-L., Teng, K. W. W., et al. (2018). Bystander CD8(+) T cells are abundant and phenotypically distinct in human tumour infiltrates. *Nature* 557, 575–579. doi:10.1038/s41586-018-0130-2

South, A. P., Purdie, K. J., Watt, S. A., Haldenby, S., den Breems, N., Dimon, M., et al. (2014). NOTCH1 mutations occur early during cutaneous squamous cell carcinogenesis. *J. Invest. Dermatol.* 134, 2630–2638. doi:10.1038/jid.2014.154

Stratigos, A. J., Garbe, C., Dessinioti, C., Lebbe, C., Bataille, V., Bastholt, L., et al. (2020a). European interdisciplinary guideline on invasive squamous cell carcinoma of the skin: part 1. epidemiology, diagnostics and prevention. *Eur. J. Cancer* 128, 60–82. doi:10.1016/j.ejca.2020.01.007

Stratigos, A. J., Garbe, C., Dessinioti, C., Lebbe, C., Bataille, V., Bastholt, L., et al. (2020b). European interdisciplinary guideline on invasive squamous cell carcinoma of the skin: Part 2. Treatment. *Eur. J. Cancer* 128, 83–102. doi:10.1016/j.ejca.2020.01.008

Wang, S., Song, R., Wang, Z., Jing, Z., Wang, S., and Ma, J. (2018). S100A8/A9 in inflammation. *Front. Immunol.* 9, 1298. doi:10.3389/fimmu.2018.01298

Wang, T., Dai, L., Shen, S., Yang, Y., Yang, M., Yang, X., et al. (2022). Comprehensive molecular analyses of a macrophage-related gene signature with regard to prognosis, immune features, and biomarkers for immunotherapy in hepatocellular carcinoma based on WGCNA and the LASSO algorithm. *Front. Immunol.* 13, 843408. doi:10.3389/fimmu.2022.843408

Yu, G., Wang, L.-G., Han, Y., and He, Q.-Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* 16, 284–287. doi:10.1089/omi.2011.0118

Zhang, G., Yan, G., Fu, Z., Wu, Y., Wu, F., Zheng, Z., et al. (2021). Loss of retinoic acid receptor-related receptor alpha (Rora) promotes the progression of UV-induced cSCC. *Cell Death Dis.* 12, 247. doi:10.1038/s41419-021-03525-x

Zhang, Z., Westover, D., Tang, Z., Liu, Y., Sun, J., Sun, Y., et al. (2024). Wnt/β-catenin signaling in the development and therapeutic resistance of non-small cell lung cancer. *J. Transl. Med.* 22, 565. doi:10.1186/s12967-024-05380-8

Zhou, X., Xia, E., Bhandari, A., Zheng, C., Xiang, J., Guan, Y., et al. (2018). LRP4 promotes proliferation, migration, and invasion in papillary thyroid cancer. *Biochem. Biophys. Res. Commun.* 503, 257–263. doi:10.1016/j.bbrc.2018.06.012