



Screening of potential markers for vitiligo based on bioinformatics and LASSO regression and prediction of Chinese medicine

Wei Liang^{a,b,c}, Minni Huang^{a,b,c}, Yue Sun^{a,**}, Shuyu Guan^{a,b,*}

^a School of Chinese Materia Medica, Guangdong Pharmaceutical University, Guangzhou, 510006, China

^b Key Laboratory of Digital Quality Evaluation of Chinese Materia Medica of SATCM, Guangzhou, 510006, China

ARTICLE INFO

Keywords:

Bioinformatics

Vitiligo

Immune infiltration

LASSO regression analysis

Weighted gene co-expression network analysis

ABSTRACT

Objective: This study aimed to use bioinformatics techniques to screen biomarkers related to vitiligo.

Methods: Firstly, the gene expression profiles of vitiligo were obtained from the GEO database, and differentially expressed genes (DEGs) were identified. Subsequently, Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses were performed on these differentially expressed genes. Through weighted gene co-expression network analysis (WGCNA), the core genes in the module most closely related to vitiligo were identified, and an intersection analysis was conducted with the DEGs. Next, a protein-protein interaction (PPI) network analysis was carried out on the intersection genes. Key genes were further screened using Cytohubba and least absolute shrinkage and selection operator (LASSO) regression analysis, and the roles of these key genes in immune cell infiltration were explored through single-sample gene set enrichment analysis (ssGSEA). In addition, the diagnostic effectiveness of the key genes was verified by the receiver operating characteristic (ROC) curve, and drugs related to the key genes were predicted using databases. Finally, the expression levels of these key genes were verified through reverse transcription quantitative polymerase chain reaction (RT-qPCR) and Western blot experiments.

Results: A total of 667 DEGs were identified, and the enrichment analysis mainly involved cell adhesion molecules, T cell receptor signaling pathway, etc. Nineteen core genes were screened out from the five algorithms of Cytohubba, and LASSO regression analysis further determined four key genes (*IL7R*, *GZMH*, *CD3G*, and *UBD*). Immune cell infiltration analysis showed that these four key genes had high expression in immune cells. The prediction results of traditional Chinese medicine showed that 15 traditional Chinese medicines were related to the key genes. The results of RT-qPCR showed that the expressions of *IL7R*, *GZMH*, and *CD3G* were significantly upregulated (* $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$), and Western blot showed obvious expressions of *IL7R*, *GZMH*, *CD3G*, and *UBD*.

Conclusion: This study used bioinformatics methods to explore the biomarkers of vitiligo, and verified the potential of *IL7R*, *GZMH*, and *CD3G* as novel candidate genes through *in vitro* experiments. These genes may become new targets for the diagnosis, prognosis, and treatment of vitiligo.

1. Introduction

Vitiligo is a common acquired depigmentation of the skin mucosa that occurs when the body's immune system attacks skin cells that produce melanin, the body's natural pigment.¹ It is also a chronic

autoimmune disease with an estimated prevalence of 0.5%–2% of the global population.² At present, the pathogenesis of vitiligo is not clear, but it may be related to autoimmunity, oxidative stress, melanocyte self-destruction, neurochemical factors, heredity, and environment.³ Existing studies have found that adaptive immunity plays a central role

* Corresponding author. Guangdong Pharmaceutical University, No. 280, Waihuan East Road, Higher Education Mega Center, Guangzhou, 510006, Guangdong, China.

** Corresponding author. Guangdong Pharmaceutical University, No. 280, Waihuan East Road, Higher Education Mega Center, Guangzhou, 510006, Guangdong, China.

E-mail addresses: sunyuesdz@163.com (Y. Sun), gdpugsy@126.com (S. Guan).

Peer review under the responsibility of Editorial Board of Journal of Holistic Integrative Pharmacy.

^c These authors contributed equally to this manuscript and worked as co-first authors.

<https://doi.org/10.1016/j.jhip.2025.06.003>

Received 14 February 2025; Received in revised form 2 May 2025; Accepted 3 June 2025

2707-3688/© 2025 The Authors. Publishing services by Elsevier B.V. on behalf of KeAi Communications Co. Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

in the pathogenesis of vitiligo, and in recent years, the important role of innate immune cells and immune-related pathways in the pathogenesis of vitiligo has been gradually clarified. Therefore, it has become an important direction to study the occurrence and development of vitiligo from the perspective of immunity.

This study first used R packages to screen differentially expressed genes (DEGs) between normal skin samples and vitiligo skin samples from microarray data of vitiligo. After performing enrichment analysis on DEGs and intersecting them with results from Weighted Gene Co-expression Network Analysis (WGCNA), key genes influencing vitiligo pathogenesis were identified using Cytohubba and LASSO regression analysis. The relative content of 28 types of immune cells in vitiligo was evaluated via single-sample gene set enrichment analysis (ssGSEA), and the correlation between key genes and immune cell infiltration was analyzed to screen immune-related key genes. Receiver operating characteristic (ROC) curves were plotted based on the specificity of key genes to assess their diagnostic efficacy. Databases such as Coremine and BenCaozuJian were used to predict traditional Chinese medicines associated with key genes, while the L1000 FWD database was employed to predict drug components related to key targets for molecular docking. Finally, *in vitro* experiments using RT-qPCR and Western blot were conducted to verify the mRNA expression levels of these key genes. This study aims to provide a theoretical basis for exploring the molecular mechanisms underlying vitiligo development.

2. Methods

2.1. Chip data collection, processing, and screening of differentially expressed genes

Using the keyword "vitiligo" to search the public GEO database (<https://www.ncbi.nlm.nih.gov/geo/>), the vitiligo datasets GSE53146 and GSE75819 were obtained. The GSE53146 dataset, based on the GPL14951 platform, contains 5 lesional skin samples from vitiligo patients and 5 normal skin samples as controls, serving as the meta-dataset. GSE75819 was used as the validation set, comprising 15 vitiligo patient skin samples and 15 normal skin samples. Data processing was performed using R software with the limma package for analysis. This included outlier processing, filtering of low-expression genes, normalization, and log2 transformation. GEO2R analysis (utilizing the Limma and Heatmap packages in R) was applied to process the GSE53146 meta-dataset, with thresholds set at $P < 0.05$ and $|\log_2 \text{FC}| > 0.5$ to identify upregulated and downregulated genes. Relevant volcano plots and heatmaps were subsequently generated.

2.2. Functional enrichment analysis and weighted gene co-expression network analysis (WGCNA) of DEG

The "ClusterProfiler" package in R was employed to perform Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses on the identified differentially expressed genes (DEGs). A significance threshold of $P < 0.05$ was established for screening, and a pathway map illustrating the functional enrichment related to these DEGs was analyzed (Fig. 2). Additionally, WGCNA can be utilized to construct a weighted gene co-expression network, identify gene modules, discover genes that exhibit strong correlations with diseases, correlate gene modules with clinical features into a co-expression module, and pinpoint core genes within those modules.⁴ Using the WGCNA package for weighted gene co-expression network analysis, the top 25% of genes based on expression variance were selected as a screening criterion to eliminate genes with low variation. A total of 13,724 genes were used to construct the co-expression network. The dynamic tree-cutting method was employed to identify gene modules, and the module with the highest correlation ($*P < 0.05$) was selected. Finally, the genes from the highest-correlation module were intersected with the DEGs identified in the analysis.

2.3. Key genes associated with vitiligo were identified through cytoscape software and LASSO regression analysis

In the fields of cell biology and systems biology, elucidating the interactions between proteins within a cellular context through the evaluation and analysis of protein-protein interaction (PPI) networks can significantly enhance our understanding of protein functionality.^{5,6}

The STRING database (<https://cn.string-db.org/>) was utilized to analyze PPIs, and PPI networks were constructed for the intersection of WGCNA and differential genes. The minimum interaction score was set to the highest confidence level (0.400), with disconnected nodes excluded from the analysis. The resulting PPI network generated in STRING was imported into Cytoscape 3.9.3 software for further examination. Five algorithms from Cytohubba, a plugin within Cytoscape 3.9.3, were employed to rank and identify the top 20 central genes in the network, with core genes determined by selecting intersections among these five algorithms. Subsequently, LASSO regression analysis on the selected core genes was conducted using the glmnet package in R, allowing for the identification of additional critical core genes.

2.4. Analysis of immune cell infiltration

The expression of data set GSE53146 in 28 kinds of immune infiltrating cells was investigated by immune cell infiltration analysis. A single sample gene set ssGSEA algorithm and an assembled reference gene set of 782 marker genes in 28 immune cell types were used to evaluate the infiltration level of 28 immune cells based on the reference gene set.⁷ The expression profiles of the GSE53146 dataset across 28 immune cell types were analyzed, resulting in the generation of correlation heat maps and violin plots to visualize these expression patterns. Spearman correlation analysis was conducted to elucidate the relationship between key genes and immune cell infiltration, with the 'ggcorrplot' package in R employed to present the findings.

2.5. Validate the selected key genes through an external validation set

Through the public database GEO (<https://www.ncbi.nlm.nih.gov/geo/>), the dataset GSE75819 of vitiligo was obtained as an external validation set by searching with the keyword "vitiligo". The ggpubr package, ggplot2 package, and ggrepel package in R software were used to draw the expression boxplots of key genes. The expression profiles of key genes in the normal group and the vitiligo group were analyzed, and the pROC package and ggplot2 package in R software were used to draw the ROC curves to screen out the biomarkers with high diagnostic value for vitiligo. The ROC curves were drawn using the pROC package in R, and the area under the curve (AUC) value was obtained to determine the effective diagnostic value of key genes. Packages such as "ggrepel" and "ggpubr" were used for visualization to obtain boxplots to compare the differences in expression levels between the normal group and the disease group in the dataset. Using R packages, the expression levels of key genes in the normal group and the disease group were compared, and the results were visualized through heatmaps.

2.6. Prediction of potential intervention of key genes in traditional Chinese medicine

Utilizing the Coremine Guide (<https://coremine.com/medical/>) and the Medica Group database (<http://herb.ac.cn/>), we projected key genes related to traditional Chinese medicine (TCM) based on the Pharmacopoeia of the People's Republic of China (2020) and conducted a query of Chinese materia medica concerning their properties and efficacy in TCM. Radar plot maps were created to visualize the Four Natures, Five Tastes, and Meridian Tropism of herbs via the chipot online data analysis platform.⁸

2.7. Molecular docking of chemical components and the proteins of key genes

Predict potential candidate drug compounds for key genes using the online L1000 FWD (<https://maayanlab.cloud/l1000fwd/>). List the top 5 candidate drugs ranked by *P*-value. Input the obtained drug names into the PubChem (<https://pubchem.ncbi.nlm.nih.gov/>) database and download the 3D structures of the active ingredients in SDF format. Use the Open Babel 3.1.1 software to convert them into mol2 format and minimize their energy. Then, obtain the 3D structure of the receptor protein in PDB format from the RCSB PDB (<https://www.rcsb.org/>). Use the PyMOL software to remove the ligands and water molecules from the receptor protein and save it in PDB format for later use. Next, upload the prepared ligands and receptors to the AutoDockTools 1.5.6 software for molecular docking to obtain the lowest binding energy. Create a heat-map of the binding energy using R language. Finally, use the PyMOL software for visualization and plotting.

2.8. Quantitative real-time PCR (RT-qPCR) analysis was performed for validation in vitro

PIG1 and PIG3 cells (normal epidermal melanin cell line PIG1 and vitiligo melanin cell line PIG3V purchased from Shanghai Chinese Academy of Sciences) were inoculated on 6-well plates (1×10^6 /well), respectively, and cultured under 37 °C and 5% CO₂ conditions to 80% fusion. Total RNA was extracted using the RNeasy Mini Kit (Qiagen, item number: 74104), and purity was determined by NanoDrop (A260/A280 > 1.9, A260/A230 > 2.0). 1 µg RNA was reverse-transcribed into cDNA (25 °C 10 min, 37 °C 120 min, 85 °C 5 min). qPCR reaction system: SYBR Green Master Mix 10 µL, cDNA 2 µL, primers 0.3 µM, ddH₂O supplementation to 20 µL. Cycle procedure: 95 °C 10 min → (95 °C 15 s, 60 °C 60 s) × 40 cycles → Melt curve analysis: 2^{−ΔΔCt} method was used to calculate the relative gene expression of each gene and GAPDH gene. Biological repeats were repeated 3 times in each group. The primer sequences of each gene are shown in Table 1. The obtained data is drawn by using ggplot2 and dplyr packages in R.

2.9. Western blotting experiments

After the cells grow to the logarithmic phase, discard the culture medium, wash twice with pre-cooled PBS, add RIPA lysis buffer (containing protease inhibitors), and lyse on ice for 30 min. Centrifuge at 4 °C and 12,000 rpm for 15 min, collect the supernatant, and determine the protein concentration by the BCA method. Take 20 µg of the protein sample, perform SDS-PAGE electrophoresis, and then transfer it to a PVDF membrane. Block with 5% skimmed milk for 1 h. Add primary antibodies against *IL7R*, *GZMH*, *CD3G*, *UBD*, and GAPDH, respectively (dilution ratio 1:1000), and incubate at 4 °C overnight. Wash the membrane three times with TBST, add the HRP-labeled secondary antibody (1:5000), and incubate at room temperature for 1 h, then perform ECL development. Use the Image Lab software to analyze the gray value of the bands, and calculate the relative expression level of the target protein with GAPDH as the internal reference. Repeat the experiment 3 times, and the data are expressed as mean ± standard deviation. Use Graphpad Prism9.5 for the *t*-test and the preparation of the bar chart. A *P* < 0.05 indicates a statistically significant difference.

Table 1
PCR primer sequences.

Gene	Forward primer (5′-3′)	Reverse primer (5′-3′)
<i>IL7R</i>	GCTGGCTCTGGTCTCTCTAA	TGGCTGTGTTGGTGTGTCT
<i>GZMH</i>	CAGCCAGACCTTCTCAACAT	TGGTGGTGTGCTGTCTTCA
<i>CD3G</i>	GACCTGGGCTACCTGAAGAT	CTGCTGTGCTTCTGGTCTT
<i>UBD</i>	AGCCCTTCTGAAGACCTGT	GCTGCACAGTGCCTTCAGA
<i>GAPDH</i>	GGAGCGAGATCCCTCCAAAAT	GGCTGTTGTACATACTTCTCATGG

3. Results

3.1. Chip data quality control and differential gene expression results

Multiple R packages in R, such as GEOquery, tidyverse, limma, and ggfortify, were used to analyze the quality standard of the dataset GSE53146, and the relative expression values of the data were normalized, and the results of data principal component analysis were visualized. Then, two groups of data in GSE53146 were analyzed. The conditions were set as *P* < 0.05, |log2FC| > 0.5. A total of 667 differentially expressed genes were screened out and visualized by volcano map, in which 320 genes were up-regulated and 347 genes were down-regulated. The top 25 genes with great differences were visualized by heat map, as shown in Fig. 1.

3.2. GO enrichment analysis and KEGG signal pathway results

R software was used to analyze the functions of differential genes and their potential mechanisms of action. Under condition *P* ≤ 0.05, a total of 17 GO items were obtained in the GO functional enrichment analysis, which were mainly classified as biological process (BP) and cell composition (CC). Bar charts and bubble charts were visualized for the GO analysis results (Fig. 2 A–B). Where BP entry, it mainly involves monocyte differentiation, T cell selection, thymus T cell selection, lymphocyte differentiation, antigen receptor-mediated signaling pathway, T cell differentiation, immune response activation signaling pathway, T cell receptor signaling pathway, α-β T cell activation, immune response regulatory signaling pathway, immune response activated cell surface receptor signaling pathway, T cell differentiation in the thymus, immune response regulatory cell surface receptor signaling pathway, positive T cell selection, etc. CC articles mainly involve the outer plasma membrane, cytolysed particles, and α-β T cell receptor complexes. In terms of the KEGG pathway (Fig. 2C–D), these genes are mainly enriched in cell adhesion molecules, T cell receptor signaling pathways, Th7 cell differentiation, primary immunodeficiency, and the intestinal immune network that produces immunoglobulin A (IgA). WGCNA in the R package was used to eliminate genes with small differences by using the expression variance of the top 25% genes as the screening condition, and a value of 10 was selected as the best soft threshold (*R*² = 0.9) to establish a scale-free network (Fig. 3A). Subsequently, cluster analysis was used to identify highly similar modules, and gene cluster trees (Fig. 3B) and module-trait information association analysis charts (Fig. 3C) were made respectively, with the minimum module size set to 60. Ten gene modules were obtained by dynamic cross-shear method, among which the red module (265 genes) had the highest correlation with vitiligo (*cor* = 0.73; *P* = 0.02), (Fig. 3C). The genes in the red module were intersected with the differential genes obtained by analysis, and the results were shown in venn diagram (Fig. 3D), 124 intersecting genes were obtained.

3.3. Cytoscape software and LASSO regression analysis were used to identify the results of vitiligo core genes

The differential gene PPI network file constructed by STRING database was calculated through five algorithms of cytoscape plugin cytoHubba (MCC, MNC, Degree, Closeness, and Radiality), and the top 20 key genes were calculated. Then, the intersection results of the five algorithms were obtained. 19 Overlapping key genes were screened, including *CD2*, *CD27*, *IL7R*, *GZMA*, *CD3E*, *NKG7*, *CD3D*, *CD247*, *GZMH*, *CD163*, *CD96*, *CD3G*, *SELL*, *IL10RA*, *UBD*, *JAML*, *GBP5*, *KIR2DL3*, and *TBC1D10C*. The 19 hub genes obtained in R packages such as glmnet, caret, randomForest, and shape were used for LASSO regression. Four key maker genes were obtained by LASSO regression analysis, namely *IL7R*, *GZMH*, *CD3G*, and *UBD*. LASSO algorithm results were shown in Fig. 4.

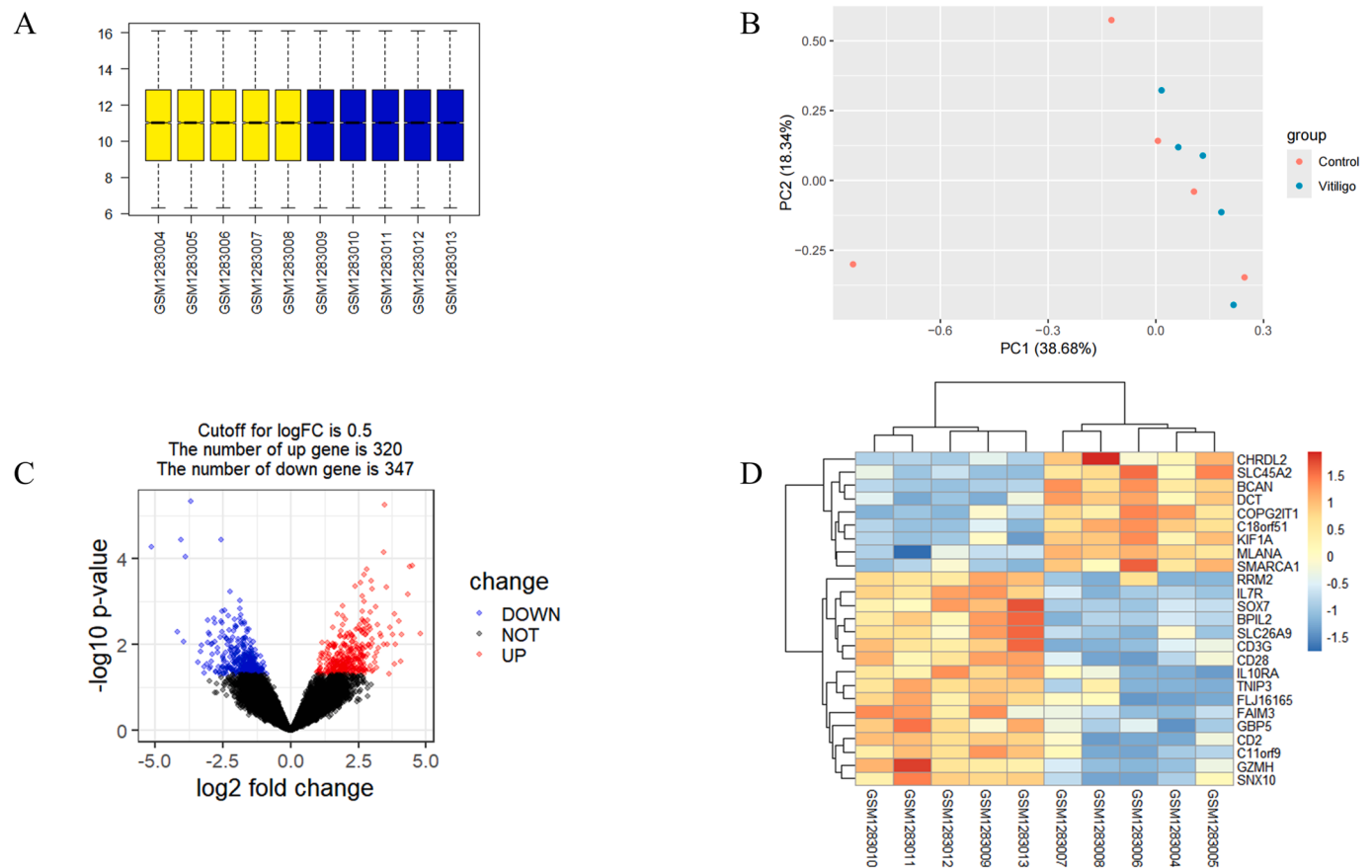


Fig. 1. Chip data evaluation results and difference analysis results. A: box map; B: PCA map; C: volcano map; D: heat map.

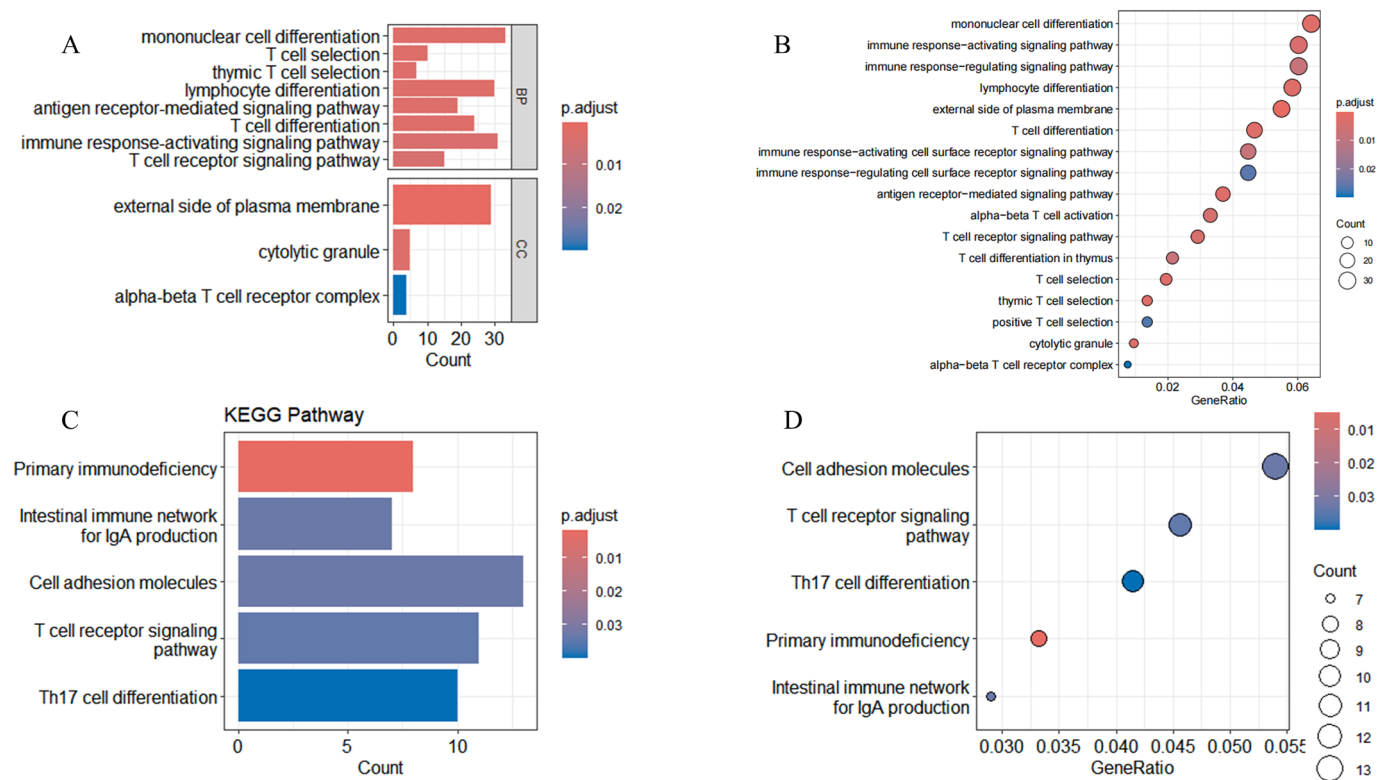


Fig. 2. GO and KEGG enrichment analysis of differentially expressed genes. A: the GO analysis bar chart; B: the GO analysis bubble chart; C: the KEGG analysis bar chart; D: the KEGG analysis bubble charts.

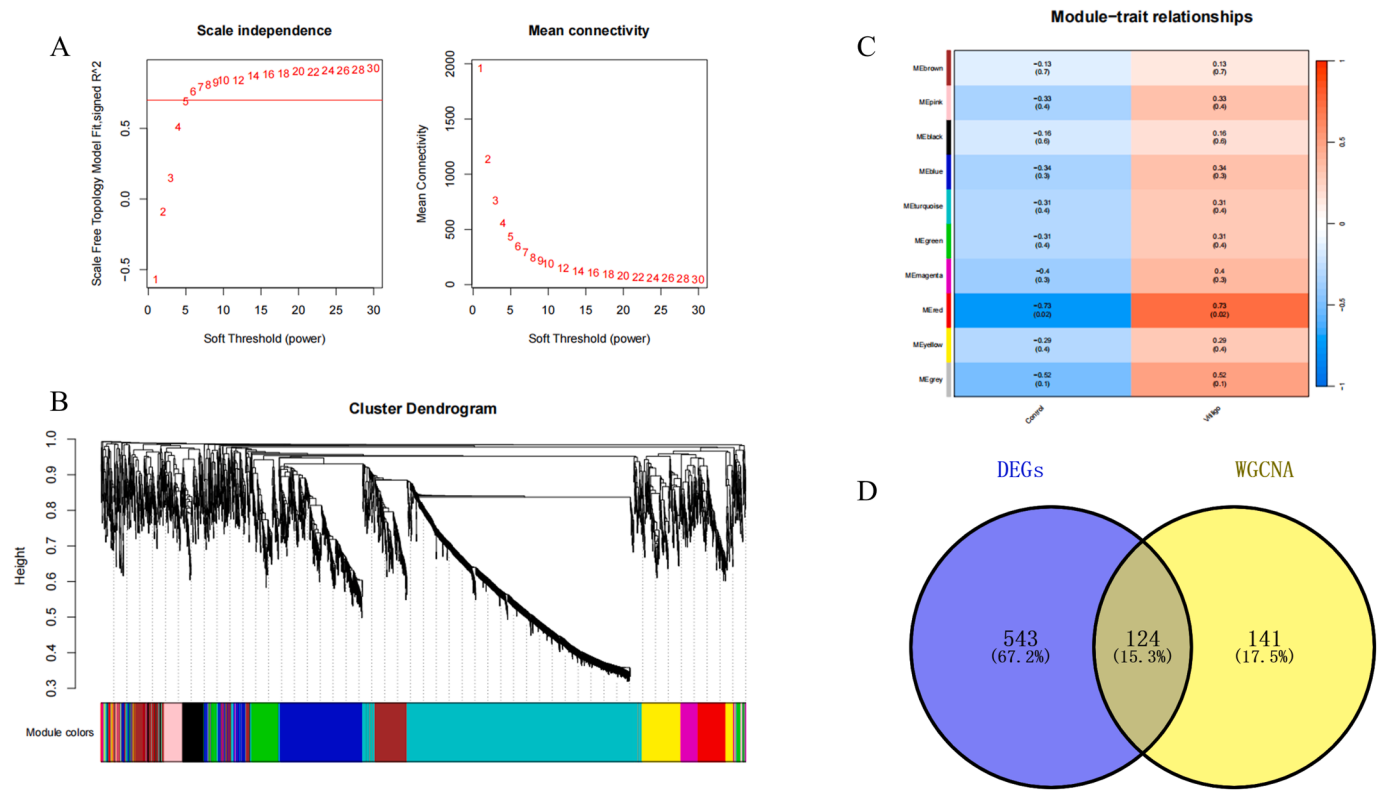


Fig. 3. WGCNA analysis results. A: soft threshold screening; B: gene cluster tree; C: module - trait information association analysis; D: intersection Wayne diagram.

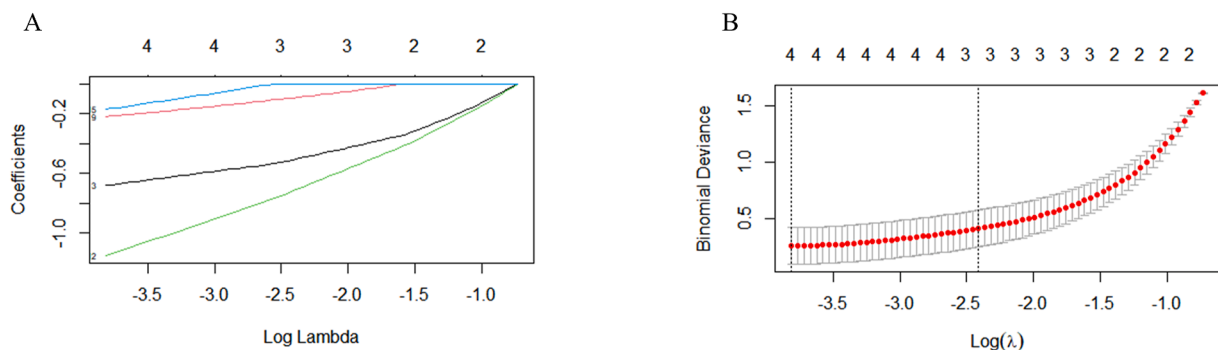


Fig. 4. Results of LASSO analysis. A is LASSO Log Lambda and B is Log(λ).

3.4. Immune cell infiltration and its relationship with diagnostic genes

Fig. 5A shows the distribution heatmap of 28 kinds of immune cells in the GSE53146 dataset. Fig. 5C, the violin map, showed significant differences between the normal group and the vitiligo group in the 28 kinds of immune cells ($*P < 0.05$, $**P < 0.01$, $***P < 0.001$). In order to further confirm the associations between 4 key genes and these immune cells, Spearman correlation analysis was performed for each key gene and 28 kinds of immune cells. Related heat Fig. 5BD indicated that the key genes *IL7R*, *GZMH*, *CD3G*, and *UBD* were highly correlated with a variety of immune infiltrating cells.

3.5. Box plot of external validation set and diagnostic validity of key genes

Box plots were used to compare the expression levels of four key genes in the external validation set GSE75819 between the normal group and vitiligo group. Compared with the normal group, the

expression of all four biomarkers was significantly upregulated in the validation set. ROC curves were drawn to validate the screening of the four key genes (Fig. 6). All four key genes had effective diagnostic value, with AUC values greater than 0.85. The AUC values of the four key genes were 0.880, 0.920, 0.920, and 0.920, respectively. The heatmap of the external validation set GSE75819 shows that the four genes all have high expression levels in the disease group (the redder the color, the higher the expression).

3.6. Prediction of potential intervention of key genes in traditional Chinese medicine

In this study, a total of 14 Chinese medicines related to key genes were predicted. The specific information of relevant Chinese medicines is shown in Table 2 below. The main Chinese medicines are warm-related drugs, and the main menstrual normalization is liver, kidney and lung channels (Fig. 7).

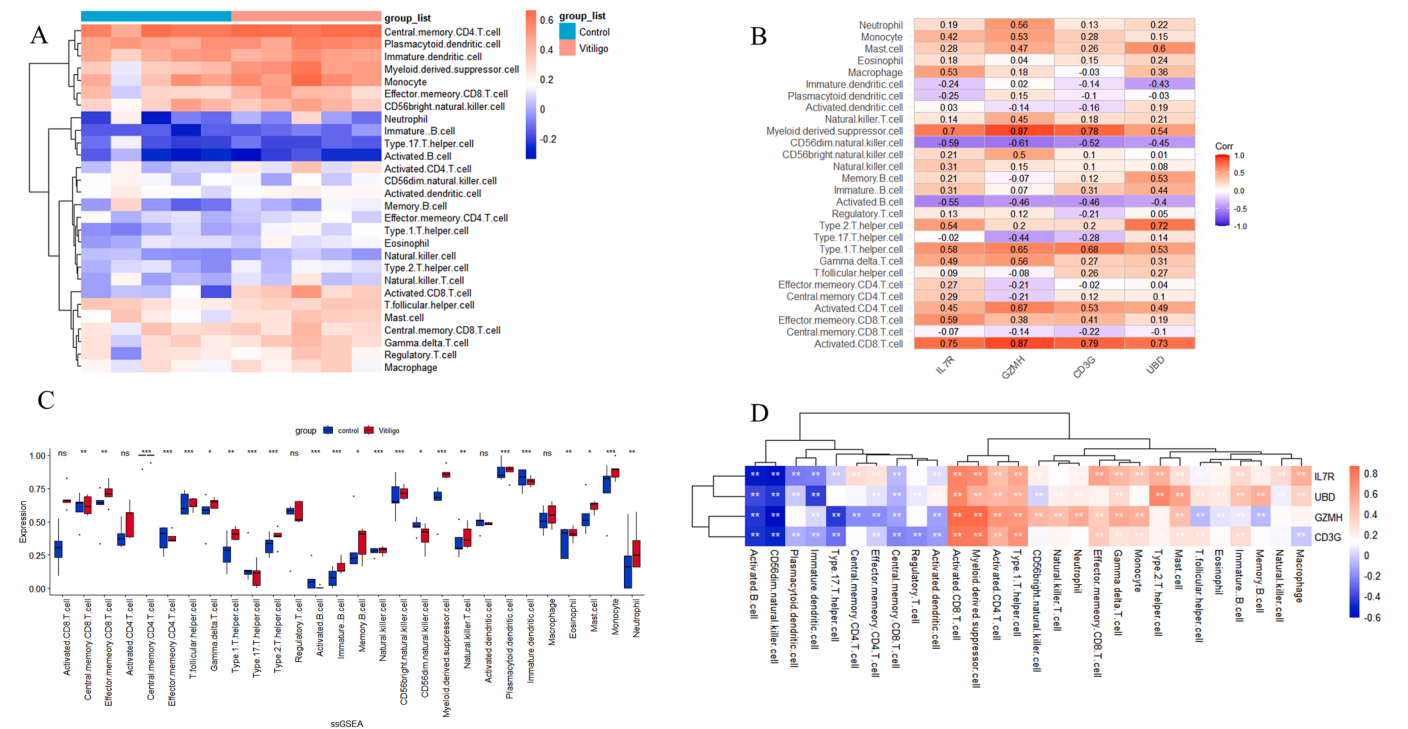


Fig. 5. Results of immune infiltration. A: the heat map of immune cells between vitiligo and normal tissue samples; B: the heat map of immune cells and 4 key genes; C: the violin map of the normal (control) group and the vitiligo group, $*P < 0.05$, $**P < 0.01$, $***P < 0.001$, compared with the control group; D: the spearman correlation heat map of immune cells and 4 key genes.

3.7. Molecular docking of chemical components and the proteins of key genes

Five top-ranked drugs were predicted, namely BRD-K58479490, sappanone-a, QL-X-138, JNJ-26854165, and CAY-10594. After performing molecular docking with key genes, the heatmap of docking binding energy is shown in Fig. 8. The lower the binding energy between the relevant compound components and key genes, the better the binding ability. The binding energy is basically above 7 kcal/mol. Then, the visualization of these components and targets was carried out (Fig. 9).

3.8. RT-qPCR analysis was performed for validation in vitro

As can be clearly seen from the bar chart in Fig. 10, the expression levels of *IL7R*, *GZMH*, and *CD3G* in vitiligo melanin cell line PIG3V were significantly higher than those of normal epidermal melanin cell line PIG1. Among them, the expression difference of *IL7R* and *CD3G* was highly significant ($P < 0.001$), and the expression difference of *GZMH* was significant ($P < 0.01$). There was no significant difference in *UBD* gene expression between the two cell lines ($P = 0.72$). This indicates that *IL7R*, *GZMH*, and *CD3G* genes may play an important role in the occurrence and development of vitiligo, and their high expression may participate in the pathophysiological process of vitiligo, while *UBD* gene may have little relationship with the pathogenesis of vitiligo.

3.9. Western blotting experiments

The results of Western blotting experiments showed that compared with the normal human epidermal melanocyte cell line PIG1, the expression levels of *IL7R*, *GZMH*, and *CD3G* proteins were significantly upregulated, while the expression level of *UBD* was significantly downregulated in the vitiligo melanocyte cell line PIG3V (Fig. 11). The loading control protein GAPDH exhibited stable expression across all groups, indicating the reliability of the experimental results.

4. Discussion

The pathogenesis of vitiligo is intricately linked to the dysregulation of the immune system. Identifying key genes that influence the immune response in vitiligo patients and elucidating their functions are not only crucial for uncovering the molecular mechanisms underlying vitiligo but also provide a foundation for early diagnosis and the development of novel therapeutic strategies.

In this study, we utilized normal skin samples and vitiligo skin samples from the GSE53146 dataset for differential expression gene analysis, identifying a total of 667 DEGs, which comprised 320 up-regulated genes and 347 down-regulated genes (Fig. 1). GO analysis of these DEGs revealed their involvement in various immune response pathways (Fig. 2A-B). Furthermore, KEGG analysis indicated that the DEGs were predominantly enriched in cell adhesion molecules, T cell receptor signaling pathways, Th7 cell differentiation processes, primary immunodeficiency mechanisms, and intestinal immune networks responsible for producing IgA (Fig. 2C-D). These findings provide compelling evidence that immune dysfunction is intricately linked to the pathogenesis of vitiligo.

In this study, we integrated five algorithms from the Cytoscape software (MCC, MNC, Degree, Closeness, and Radiality) with a LASSO regression model to identify four key genes closely associated with the onset and progression of vitiligo: *IL7R*, *GZMH*, *CD3G*, and *UBD*. To further elucidate the role of immune infiltration in vitiligo pathogenesis, we employed the ssGSEA algorithm for immune infiltration analysis on the GEO dataset. The immunoinfiltration heatmap (Fig. 5A) indicated that the infiltration ratios of central memory CD4 T cells, plasmacytoid dendritic cells, immature dendritic cells, myeloid suppressor cells, and activated B cells were significantly elevated in the vitiligo group. Additionally, the violin plot comparing immune infiltrating cell populations between normal and vitiligo tissues (Fig. 5C), revealed that compared to normal tissue samples, there was a marked increase in infiltrating type 2 T helper cells ($P < 0.001$), immature B cells ($P < 0.001$), myeloid suppressor cells ($P < 0.001$), and monocytes ($P < 0.001$) within vitiligo tissues.

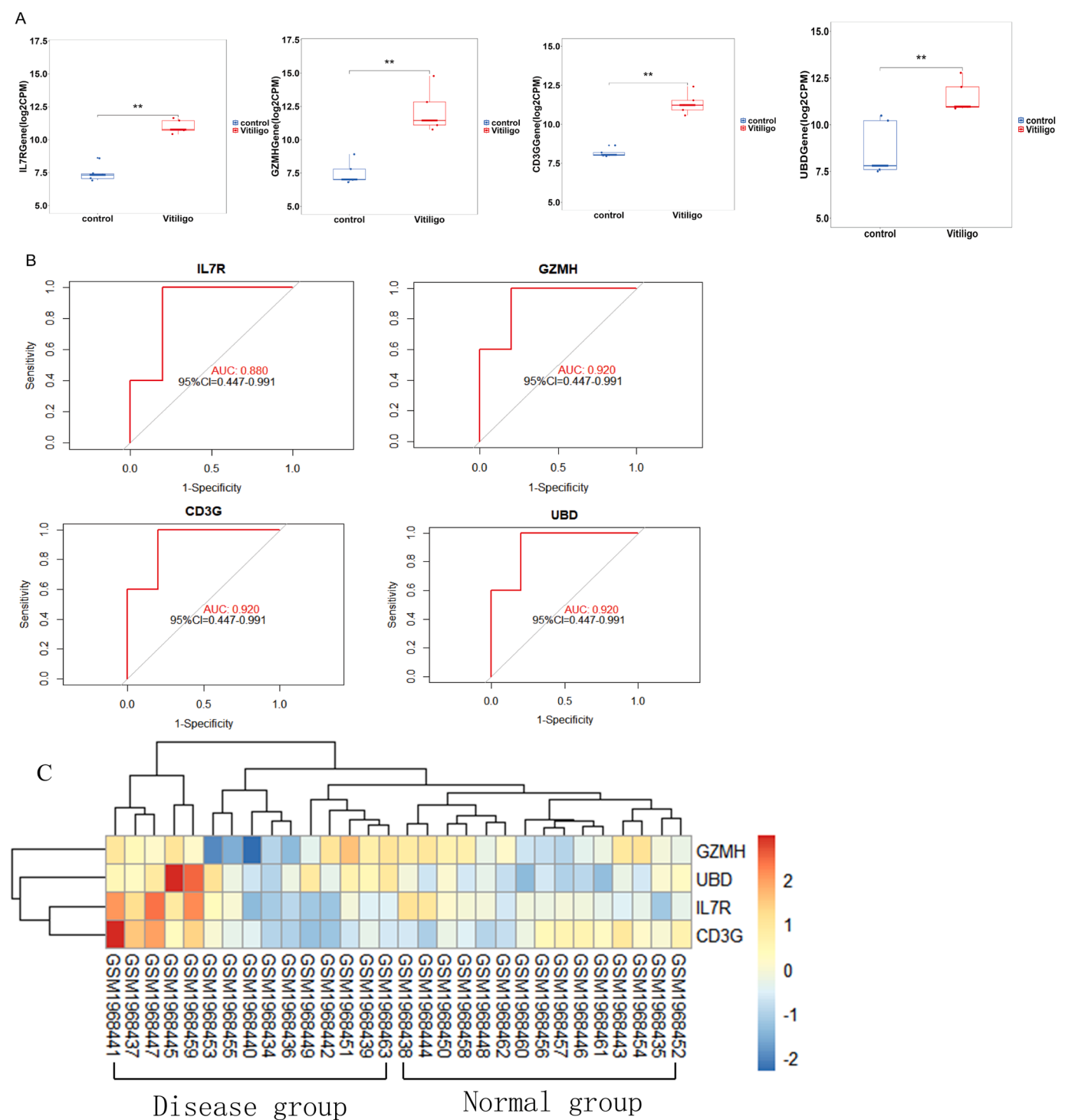


Fig. 6. Diagnostic effectiveness of key genes. A: normal group and vitiligo group box type; B: key gene ROC curve; C: Co-expression heatmap.

IL7R is a gene expressing a unique alpha chain of the interleukin-7 (IL-7) heterodimer receptor that is essential for T and B cell development, juvenile T cell survival, and memory T cell maintenance.⁹ Flow cytometry and single-cell RNA sequencing of bone marrow samples from healthy controls and *IL-7Rα* deficient patients, combined with *in vitro* models of human B cell differentiation, demonstrated that *IL-7R* signaling plays a crucial role in human B lymphocyte generation.¹⁰ The study used the CRISPR/Cas9 system to knock out the interleukin-7 receptor (*IL7R*) in zebrafish and found that the lack of *IL7R*, an essential molecule for maintaining normal retinal development in zebrafish, hindered the development of the retina.¹¹ The protein encoded by the

GZMH gene is cytotoxic granzyme, a family member of which *GZMH* belongs to. This protein is reported to be constitutively expressed in natural killer (NK) cells and can act as a cytotoxic arm of innate immune response by inducing target cell death and by direct lysis of pathogen-infected cell substrates.¹² A randomized controlled clinical trial indirectly demonstrated a positive association between *GZMH* and inflammation.¹³ The CD3G gene encodes the CD3γ subunit of the T-cell receptor complex. Dysfunction of this gene is associated with Immunodeficiency 17 and immune dysregulation disease with immunodeficiency, and it is involved in processes such as the Adenosine A2b Receptor (ADORA2B) mediated anti-inflammatory pathway and the

Table 2
Traditional Chinese medicine prediction of immune key genes in the pathogenesis of vitiligo.

Serial number	Chinese medicines	Latin name of Chinese medicines ^a	Corresponding genes	Nature and taste	Channel tropism	Function and indications
1	tie pi shi hu	Dendrobii Officinalis Caulis	IL7R	Sweet, slightly cold	Stomach, kidney	Benefiting stomach and promoting fluid, nourishing Yin and clearing heat
2	yan jing she	Naja atra	IL7R	Sweet, salty, warm	Liver, kidney	Channel and collaterals, rheumatism
3	huang qi	Astragali Radix	IL7R	Sweet, warm	Lung, spleen	Tonifying Qi, rising Yang, strengthening the surface to stop perspiration, reducing swelling, promoting fluid and blood, alleviating stagnation and impediment, expelling toxin and draining pus, restraining sores and producing muscle
4	yu jin	Curcumae Radix	IL7R	Pungent, bitter, cold	Liver, heart, lung	Invigorating blood to relieve pain, moving Qi to dissolve stagnation, clearing heart heat and cooling blood, facilitating bile flow to reduce jaundice
5	she xiang	Moschus	IL7R	Pungent, warm	Heart, spleen	Inducing resuscitation and awakening the spirit, activating blood and dredging meridians, dispersing swelling and alleviating pain
6	jiang huang	Curcumae Longae Rhizoma	IL7R	Pungent, bitter, warm	Spleen, liver	Breaking blood stasis and moving Qi, dredging meridians to alleviate pain
7	yu mi xu	Zea mays L.	UBD	Sweet, light, flat	Bladder, liver, gallbladder	Promote urination and reduce edema; clear heat; calm the liver; promote bile flow.
8	bie jia jiao	Trionyx sinensis Wiegmann	CD3G	Salty, cold	Lung, liver, kidney	Nourishing Yin, tonify blood; clear heat; resolve stasis
9	bie tou	Trionyx sinensis Wiegmann	CD3G	Sweet, slightly astringent	Spleen, large, intestine	Prolonged dysentery, postpartum uterine ptosis, vaginal sores
10	yuan can e	Bombyx mori L.	CD3G	Salty, warm	Stomach, spleen, liver	Tonifying kidney and strengthening Yang, astringent essence, hemostasis, detoxification and detumescence
11	bie jia	Trionycis Carapax	CD3G	Salty, cold	Liver, kidney	Nourishing Yin and latent Yang, reducing heat and removing steam, soft and hard
12	can sha	Bombyx mori Linnaeus	CD3G	Pungent, sweet, warm	Stomach, spleen, liver	Dispelling wind dehumidification, and stomach turbid, activating blood through menstruation
13	gui zhi	Cinnamomi Ramulus	GZMH	Pungent, sweet, warm	Heart, lungs, bladder	Assisting yang to transform Qi, pacifying upward surge and directing Qi downward
14	ling zhi	Ganoderma	GZMH	Sweet, flat	Heart, lungs, liver, kidneys	Tonifying Qi to tranquilize the spirit, arresting cough and calming panting

^a Based on Chinese Pharmacopoeia (2020) or <https://species.sciencereading.cn/home>.

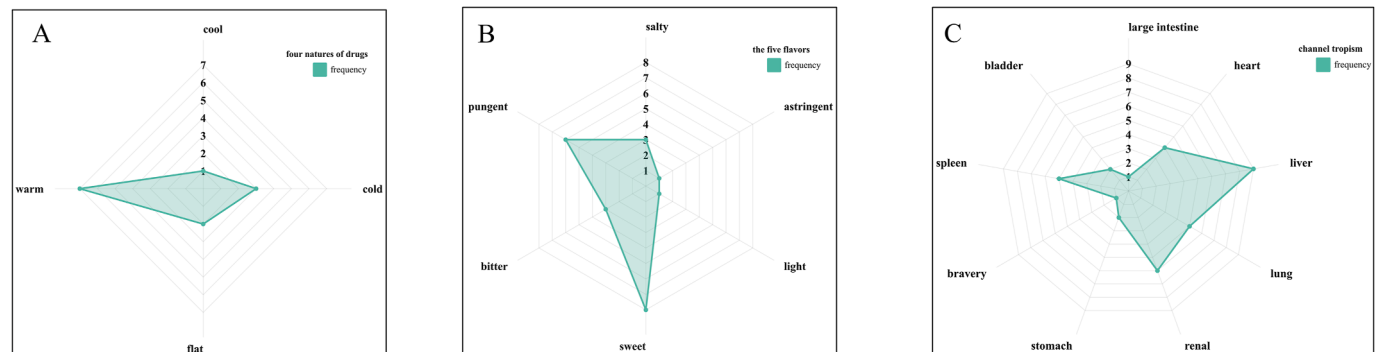


Fig. 7. Radar chart of the four natures, five flavors, and channel tropism of traditional Chinese medicine was predicted. A: four natures of drugs; B: the five flavors; C: channel.

regulation of actin dynamics during phagocytic cup formation.¹⁴ Studies have shown that *CD3G* deficient patients have decreased T cell diversity, decreased inhibitory function, and increased autoimmune clonality.¹⁵ *UBD* is a member of the ubiquitin-like protein family. It is mainly expressed in thymus, lymph nodes, and other tissues and organs of the immune system, and plays an important role in immune regulation such as immune response, antigen presentation and antiviral infection.¹⁶ Modern studies have confirmed that vitiligo is a loss of pigment caused by $CD8^+$ T cells attacking melanocytes through skin explants.¹⁷ The four key genes identified in this study exhibit elevated expression levels in $CD8^+$ T cells, potentially influencing the pathogenesis of vitiligo through various immune mechanisms. The abnormal expression of the four key genes screened in this study was found to have a certain correlation with immune cell infiltration (Fig. 5B–D), which provides reference significance for in-depth research on the occurrence and development of

vitiligo. Through the combination of traditional Chinese and Western medicine prediction, potential active ingredients associated with these four genes were explored, and 14 traditional Chinese medicines were identified (Table 2). The results of ROC detection confirmed that all four key genes had good diagnostic efficacy (Fig. 6), further verifying the accuracy of the key genes. Finally, *in vitro* experiments validated by RT-qPCR showed that the genes *IL7R*, *GZMH*, and *CD3G* had high expression levels in vitiligo tissues, suggesting they may play important roles in the occurrence and development of vitiligo, with their high expression potentially involved in the pathophysiological processes of vitiligo. Although the *UBD* gene showed no significant difference in RT-qPCR but was significant in Western blotting, this discrepancy may be attributed to differences in the detection targets (protein vs. mRNA), technical principles (detection basis, sensitivity, etc.), as well as variations in sample processing, experimental repetition, and operations.

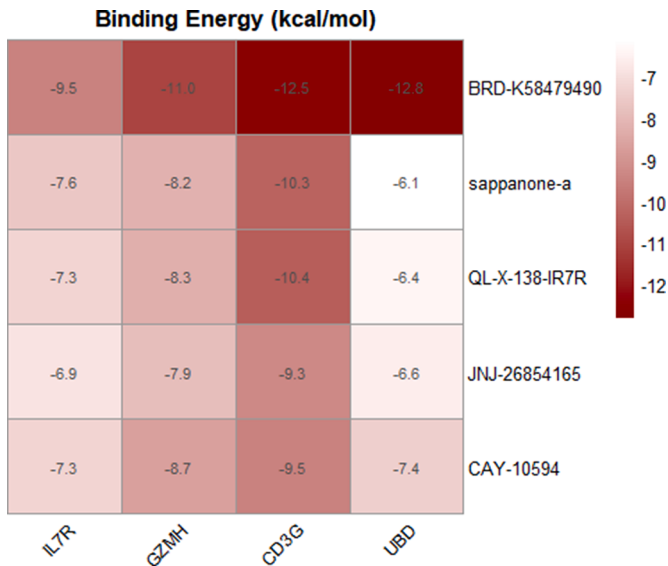


Fig. 8. Heatmap of binding energy. The redder, the higher of the docking binding energy.

Notably, *IL7R*, *GZMH*, and *CD3G* showed high expression in both experiments, further confirming their critical roles in the development of vitiligo.

Modern medicine believes that the occurrence of vitiligo is the result of the interaction of heredity, autoimmunity, melanocyte self-

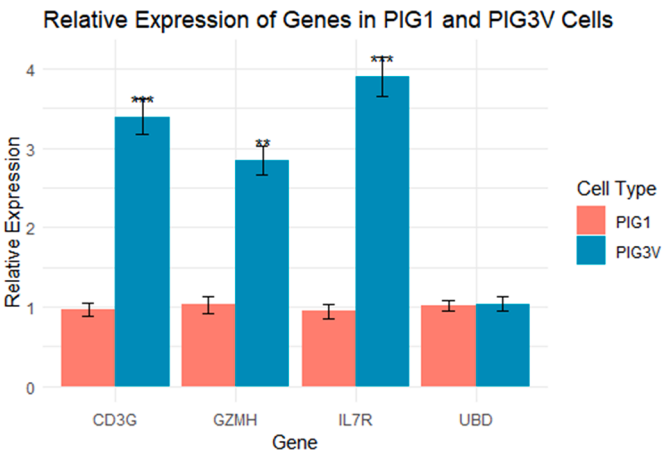


Fig. 10. mRNA expression levels bar graph. (Compared with the control group (PIG1 cell group), * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$).

destruction, and oxidative stress.³ Traditional Chinese medicine believes that the occurrence of vitiligo is closely related to the dysfunction of zang-fu organs and the six external feelings. The Taiping Sheng Hui Fang states: "Vitiligo manifests as roughened skin with white patches, arising from heat stagnation in the lung organ, invasion by wind pathogens, and the interplay of wind-heat evils. These pathogenic factors circulate through the Ying-Wei (nutritive and defensive systems), stagnate in the muscles, and persist unresolved, ultimately forming the condition." Modern clinical research in TCM highlights the critical role of the smooth regulation of liver Qi in the pathogenesis of vitiligo. The

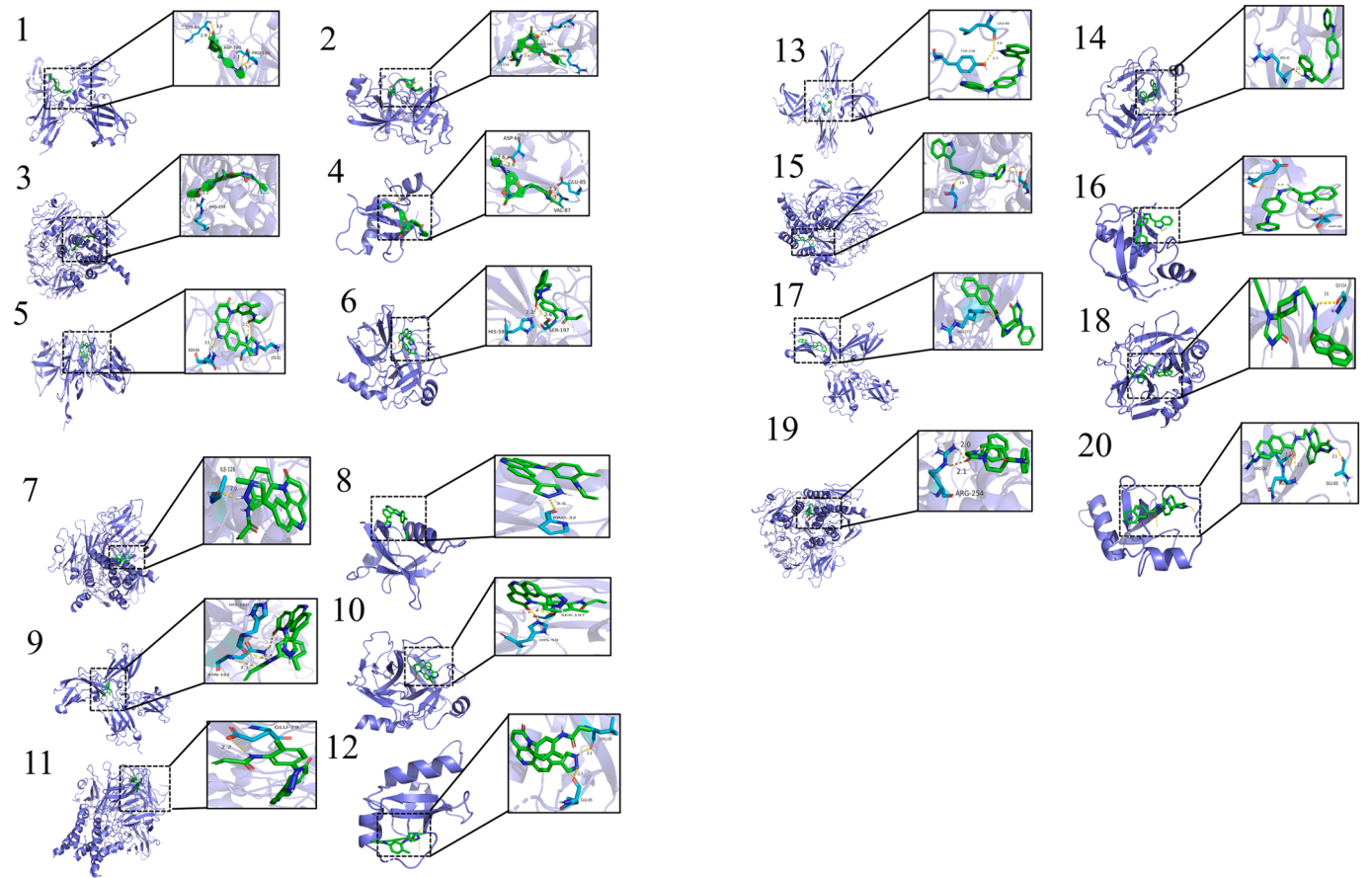


Fig. 9. Visual Results of Molecular Docking. 1–4: BRD-K58479490 with *IL7R*, *GZMH*, *CD3G*, *UBD*; 5–8: sappanone-a with *IL7R*, *GZMH*, *CD3G*, *UBD*; 9–12: QL-X-138 with *IL7R*, *GZMH*, *CD3G*, *UBD*; 13–16: JNJ-26854165 with *IL7R*, *GZMH*, *CD3G*, *UBD*; 17–20: CAY-10594 with *IL7R*, *GZMH*, *CD3G*, *UBD*.

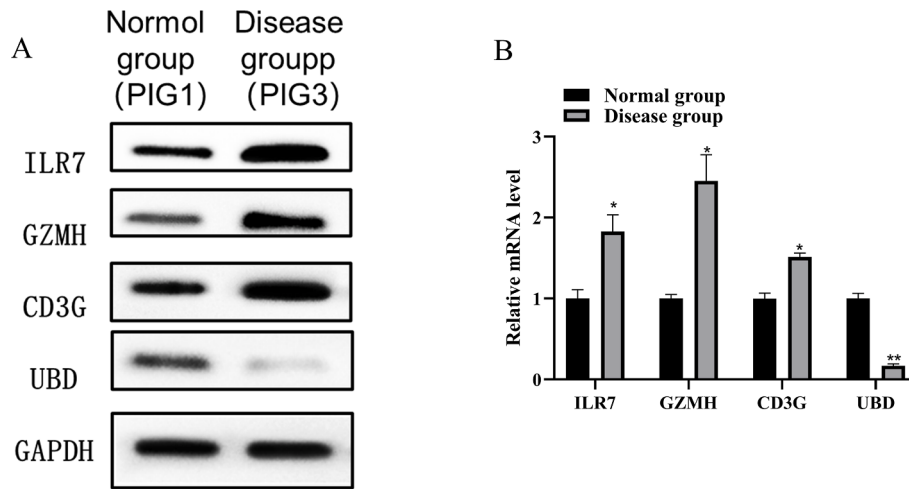


Fig. 11. Western blot detected the expression levels of IL7R, GZMH, CD3G, and UBD in PIG1 and PIG3V cells, respectively. A. The results of a Western blot; B. The relative expression levels of the four key gene proteins. (Compared with the Normol group (PIG1 cell group), * $P < 0.05$, ** $P < 0.01$).

characteristic white patches align with TCM theory: the lung governs the color white (one of the five elemental correspondences), while the kidney governs black. The lung (metal) and kidney (water) are mutually nourishing (metal-water interaction), and deficiency in these organs may disrupt this relationship, contributing to the development of white patches.^{18,19} In this study, the traditional Chinese medicine related to key genes was searched through databases, and its four natures, five flavors, and channel tropism were analyzed. The predicted results were consistent with the TCM syndrome differentiation. Therefore, the predicted Chinese medicine may be a potential medicine for the treatment of vitiligo.

5. Conclusion

Through integrated bioinformatics analysis and experimental validation, this study systematically identified immune-related biomarkers in vitiligo. Using GEO datasets and applying differential gene expression analysis, WGCNA, and LASSO regression, four key genes (*IL7R*, *GZMH*, *CD3G*, and *UBD*) were screened as core participants in the pathogenesis of vitiligo. Functional enrichment analyses indicated that these genes are primarily involved in immune pathways, particularly T-cell receptor signaling and immune cell infiltration. Immune infiltration analysis further revealed significant correlations between these genes and elevated populations of immune cells such as CD8⁺ T cells and myeloid-derived suppressor cells, highlighting their roles in driving autoimmune responses.

Prediction analysis of traditional Chinese and Western medicines identified 14 traditional Chinese medicines associated with these key genes. Validation via RT-qPCR and Western blot experiments confirmed the significant upregulation of *IL7R*, *GZMH*, and *CD3G* in vitiligo melanocytes, supporting their potential as diagnostic or therapeutic targets. Explorations of predicted traditional Chinese medicines and molecular docking studies of Western drugs provided new insights into potential interventions targeting these genes, aligning with traditional Chinese medicine theories of immune and organ system regulation.

Although discrepancies in significance for *UBD* were observed between the two experimental validations, the robust diagnostic efficacy (AUC > 0.85) and immune-related functions of the other three genes underscore their translational value. Future research should focus on elucidating the mechanistic roles of these genes in melanocyte destruction and validating the predicted herbal compounds in preclinical models. This study establishes a foundation for advancing precision medicine in vitiligo, bridging bioinformatics discoveries with experimental and traditional therapeutic strategies.

6. Limitations

This study is constrained by the limited sample size of the meta-dataset (GSE53146) and potential batch effects in cross-platform validation, methodological dependencies on parameter-driven algorithms (e.g., WGCNA, LASSO) that may affect reproducibility, incomplete experimental validation of UBD and lack of in vivo models, unverified predictive outcomes for traditional Chinese medicines and molecular docking interactions, and insufficient incorporation of clinical variables or non-immune pathways (e.g., oxidative stress) that might synergistically contribute to vitiligo pathogenesis.

CRedit authorship contribution statement

Wei Liang: Writing – review & editing. **Minni Huang:** Writing – original draft. **Yue Sun:** Supervision. **Shuyu Guan:** Resources.

Data availability

Data will be made available on request.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by Research on Policies of Traditional Chinese Medicine in the Context of Deepening Healthcare Reform from the Monitoring and Statistical Research Center of the National Administration of Traditional Chinese Medicine, China (grant numbers YGZXKT2024313).

References

- Lewitt TM, Kundu RV. Vitiligo. *JAMA Dermatol.* 2021;157(9):1136.
- Bergqvist C, Ezzedine K. Vitiligo: a review. *Dermatology.* 2020;236(6):571–592.
- Zhang C, Long J. Research progress on immune mechanism of vitiligo and intervention of traditional Chinese medicine. *Yunnan J Tradit Chin Mater Med.* 2023;44(12):79–82.
- Chen S, Yang D, Lei C, et al. Identification of crucial genes in abdominal aortic aneurysm by WGCNA. *Peer J.* 2019;7:e7873.
- Li M, Lu Y, Niu Z, et al. United complex centrality for identification of essential proteins from PPI networks. *IEEE ACM Trans Comput Biol Bioinf.* 2017;14(2):370–380.

6. Manuel MF, Miguel AVR. Decomposition-based multi-objective optimization approach for PPI network alignment. *Knowl-Based Syst.* 2022;243(11):108527.
7. Zhang B, Wang Z. A novel pyroptosis-regulated gene signature for predicting prognosis and immunotherapy response in hepatocellular carcinoma. *Front Mol Biosci.* 2022;9:890215.
8. Ji X, Tang J, Zhang J. Effects of salt stress on the morphology, growth and physiological parameters of *Juglans microcarpa* L. seedlings. *Plants (Basel).* 2022;11(18):2381.
9. Wang X, Chang S, Wang T, et al. IL7R is correlated with immune cell infiltration in the tumor microenvironment of lung adenocarcinoma. *Front Pharmacol.* 2022;13:857289.
10. Kaiser FMP, Janowska I, Menafrá R, et al. IL-7 receptor signaling drives human B-cell progenitor differentiation and expansion. *Blood.* 2023;142(13):1113–1130.
11. Cai S, Chen Y, Shang Y, et al. Knockout of zebrafish interleukin 7 receptor (IL7R) by the CRISPR/Cas9 system delays retinal neurodevelopment. *Cell Death Dis.* 2018;9(3):273.
12. Lenárt M, Bober P, Marcin M, et al. Peripheral blood CD8⁺ T-lymphocyte immune response in benign and subpopulations of breast cancer patients. *Int J Mol Sci.* 2024;25(12):6423.
13. Sun G, Zhou Y, Han X, et al. Potential marker genes for chronic obstructive pulmonary disease revealed based on single-cell sequencing and Mendelian randomization analysis. *Aging (Albany NY).* 2024;16(10):8922–8943.
14. Gene Cards. *CD3G Gene - CD3 Gamma Subunit Of T-Cell Receptor Complex*. Account - genecards suite; 2025. March 8, 2025. Accessed March 17, 2025 [https:// www.genecards.org/cgi-bin/carddisp.pl?gene= CD3G](https://www.genecards.org/cgi-bin/carddisp.pl?gene=CD3G).
15. Gokturk B, Keles S, Kirac M, et al. CD3G gene defects in familial autoimmune thyroiditis. *Scand J Immunol.* 2014;80(5):354–361.
16. Liu Y, Cheng K, Sun M, et al. UBD participates in neutrophilic asthma by promoting the activation of IL-17 signaling. *Int J Biol Macromol.* 2024;264(Pt1):130581.
17. Vu TT, Koguchi-Yoshioka H, Watanabe R. Skin-resident memory T cells: pathogenesis and implication for the treatment of psoriasis. *J Clin Med.* 2021;10(17):3822.
18. Wang P, Jia T, Zhao Y, et al. Study on differentiation and treatment of vitiligo in Chinese medicine. *Chin J Dermatol.* 2020;19(2):187–189.
19. Zhang X, Meng L, Li S, et al. Pathogenesis of vitiligo and progress of TCM treatment. *J Guizhou Univ TCM.* 2024;46(3):57–63.