



OPEN ACCESS

EDITED BY

Katarzyna Wojciechowska-Durczynska,
Medical University of Lodz, Poland

REVIEWED BY

Yong Cui,
China-Japan Friendship Hospital, China
Katarzyna Poznańska-Kurowska,
Bieganski Hospital, Poland

*CORRESPONDENCE

Yan Wu

✉ wuyan2001@163.com

RECEIVED 11 December 2024

ACCEPTED 21 April 2025

PUBLISHED 15 May 2025

CITATION

Wang Q, Yuan J, Zhang M, Jia H, Lu H and
Wu Y (2025) Bioinformatics meets machine
learning: identifying circulating biomarkers for
vitiligo across blood and tissues.
Front. Immunol. 16:1543355.
doi: 10.3389/fimmu.2025.1543355

COPYRIGHT

© 2025 Wang, Yuan, Zhang, Jia, Lu and Wu.
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.

Bioinformatics meets machine learning: identifying circulating biomarkers for vitiligo across blood and tissues

Qiyu Wang¹, Jingwei Yuan^{1,2}, Mengdi Zhang², Haiyan Jia²,
Hongjie Lu² and Yan Wu^{2*}

¹Beijing Technology and Business University, Beijing, China, ²Air Force Medical Center of the Chinese People's Liberation Army, Beijing, China

Background: Vitiligo is a skin disorder characterized by the progressive loss of pigmentation in the skin and mucous membranes. The exact aetiology and pathogenesis of vitiligo remain incompletely understood.

Methods: First, a microarray dataset of blood samples from multiple patients with vitiligo was collected from GEO database. The limma package was used to analyze the microarray data and identify significant differentially expressed genes (DEGs). The merged microarray data were then used for WGCNA to identify modules of features genes. DEGs selected with the limma package and module genes derived from the WGCNA were intersected using the Venn package in R. Enrichment analyses were performed on the overlapping genes, including Gene Ontology and Kyoto Encyclopedia of Genes and Genomes methodology. Advanced screening was performed using the least absolute shrinkage and selection operator and support vector machine techniques from the machine learning toolkit. CIBERSORT was used to analyse the immune cell composition in the microarray data to assess the relationships among these genes and immune cells. Biological samples were obtained from the patients, and gene expression analysis was performed to evaluate the levels of core genes throughout the progression of vitiligo. Finally, we obtained the microarray datasets GSE53146 and GSE75819 from the affected skin of vitiligo patients and GSE205155 from healthy skin to perform expression analysis and gene set enrichment analysis of the hub genes.

Results: Two hub genes, *HMGA1* and *PSMD13*, were identified via machine learning and WGCNA. The analysis of immune cell infiltration suggested that different immune cell types could play a role in the progression of vitiligo. Moreover, these hub genes exhibited varying degrees of association with immune cell profiles. qRT-PCR analysis of blood samples from vitiligo patients revealed notable downregulation of the hub genes. Analysis of the microarray datasets derived from skin lesions revealed that *HMGA1* expression levels remained relatively stable, whereas *PSMD13* expression levels markedly decreased.

Conclusion: *PSMD13* may influence vitiligo development via the Nod-like receptor signaling pathway and could serve as a potential diagnostic marker for evaluating skin lesions in vitiligo.

KEYWORDS

vitiligo, pigmentation loss, WGCNA, machine learning, immune infiltration, biomarkers

1 Introduction

Vitiligo is an autoimmune dermatological disorder characterized by the loss of melanocytes, the cells that produce skin pigments, leading to areas of depigmentation that manifest as white patches on the skin (1, 2). Statistics indicate that the global incidence rate is 0.36%. In adults, the rate is around 0.67%, while in children it is about 0.24%. Overall, approximately 28.5 million people are affected (3). In addition, due to its negative impact on patients' social interactions, vitiligo leads to a high prevalence of mental disorders such as anxiety and depression (4, 5), which leads to a significant decline in the quality of life of patients.

The pathogenesis of this disease has not been fully elucidated, and current research has focused primarily on oxidative stress promoting the CD8+ T-cell response in the skin (6, 7). Inherent lymphocytes mediate the apoptosis of melanocytes (8) and the effects of family genetics (9). Present treatments include ultraviolet light-based combination therapies (10), targeted T-cell therapies (11) and JAK inhibitors, but the likelihood of achieving long-term colour recovery is quite small (12). Therefore, investigating the mechanisms underlying vitiligo, identifying biomarkers for this disease, and developing prevention and treatment strategies on the basis of these findings are highly important.

In recent years, advancements in bioinformatics have led to its widespread application along with machine learning techniques for identifying treatment targets for various diseases (13). Gene chips, a high-sequence technique for analysing RNA expression, have been used in medical research into numerous disorders, including neurodevelopmental disorders (NDDs) (14).

Weighted gene co-expression network analysis (WGCNA) is a powerful method for analysing transcriptomic data. In WGCNA, tightly connected genes are clustered into different modules, not only identifying differentially expressed genes (DEGs) but also exploring the correlations between modules and diseases (15). The identification of disease biomarkers through scale-free WGCNA has been utilized in numerous disease studies (16–18). Machine learning encompasses a range of mathematical techniques designed to extract insights from large datasets (19). Machine learning algorithms offer several advantages, such as non-linearity, fault tolerance, and real-time operation, that make them well suited for complex applications (20).

In this study, we obtained two expression microarray datasets of peripheral blood mononuclear cells (PBMCs) from vitiligo patients

and healthy controls from the Gene Expression Omnibus (GEO) database. Differential gene identification ($n=154$) was conducted with the limma and WGCNA packages in R. Machine learning algorithms were used to select two core genes. Additionally, we assessed the proportions of 22 immune cell types in the PBMCs of vitiligo patients and healthy individuals via the CIBERSORT algorithm. Biological samples were collected, and the expression levels of these core genes throughout vitiligo pathogenesis were evaluated. Finally, we collected microarray datasets from lesional skin and healthy skin for expression analysis and gene set enrichment analysis (GSEA) of the hub genes. The analysis process of this study is illustrated in Figure 1.

2 Materials and methods

2.1 Gene expression profile (dataset download and processing)

Data were collected from the GEO database (<https://www.ncbi.nlm.nih.gov/geo/>), a public resource for high-throughput gene expression data, chips, and microarrays. We obtained two datasets derived from blood samples from vitiligo patients: GSE90880 (GPL8300), which included nine vitiligo patients and six healthy controls, and GSE80009 (GPL16951), which included eight vitiligo patients and four healthy controls. Three datasets derived from skin samples of vitiligo patients and healthy controls were used for data validation: GSE53146 (GPL14951), which included 5 vitiligo patients and 5 healthy controls; GSE75819 (GPL6884), which included 15 vitiligo patients; and GSE205155 (GPL14550), which included 11 healthy controls.

2.2 Identification of DEGs

We compared vitiligo subjects with healthy control subjects using R (v4.2.2). The R package “limma” was used to identify DEGs between the vitiligo patient group and healthy control group for both the GSE90880 and GSE80009 datasets. Adjusted p values were evaluated to mitigate the risk of false positives in the GEO dataset. DEGs with adjusted p values < 0.05 and $|\log_2 \text{fold change} (\log_2 \text{FC})| > 0.5$ were deemed significant. Volcano maps were created with a

volcano plotting tool (<http://soft.sangerbox.com/>). Heatmaps based on filtered DEGs were generated with the pheatmap package in R.

2.3 Weighted gene co-expression networks

The gene expression profile was generated using the “sva” package to normalize the data and correct for batch effects. The GSE90880 and GSE80009 datasets were subsequently merged. The Pearson correlation coefficient was computed with the R WGCNA software package to determine the correlation between gene pairs, establishing the gene co-expression matrix. Following the scale-free network principle, soft thresholds ($\text{power} = 12$, $R^2 = 0.89$) were chosen successively to construct scale-free co-representation networks. The adjacency matrix was then converted into a topological overlap matrix. Subsequently, cluster analysis was performed to delineate gene modules, each containing at least 60 genes. Hierarchical clustering was employed to generate a dendrogram, evaluating the correlation between module feature genes and the disease phenotype. The module with the highest correlation coefficient and the lowest P value was identified as the disease feature. Next, the module feature genes obtained using limma were intersected with the GSE90880 and GSE80009 datasets, and a Venn diagram was generated to identify the hub genes.

2.4 Selection of feature genes for Gene Ontology and Kyoto encyclopedia of genes and genomes functional enrichment analyses

The intersecting genes identified by limma and WGCNA were selected using the Venn software package in R.

In this study, the “clusterProfiler” R package was used to perform GO and KEGG functional enrichment analyses, with a focus on gene-related biological process (BP), molecular function (MF), and cellular component (CC) terms and their associated signalling pathways.

2.5 Machine learning analysis of disease genes

We employed three machine learning algorithms, namely, least absolute shrinkage and selection operator (LASSO) regression analysis, support vector machine with recursive feature elimination (SVM-RFE) analysis, and random forest (RF) analysis, to analyse the genes identified by the aforementioned methods. We used performed the LASSO regression analysis with the R package “rms”, SVM-RFE analysis using the R package “e1071”, and RF analysis with the R package “randomForest.” The intersecting genes obtained from these three analyses are considered the core genes of vitiligo. Finally, we implemented

stratified 5-fold cross-validation in R software to evaluate the logistic regression model built on core genes. Each fold preserved the original class distribution of the dataset. After training the model on four folds, we generated predictions for the held-out test fold and merged all test fold predictions to compute the pooled AUC. To evaluate stability, we conducted 1000 bootstrap resampling on the combined predictions, reporting the mean AUC and the 95% confidence interval.

2.6 Immune infiltration analysis

CIBERSORT is a computational tool that was used to estimate the proportions of immune cells in vitiligo and control samples on the basis of tissue gene expression profiles, helping to identify variations in immune cell composition. We analysed immune cell infiltration with the “Cibersort” R software package. Bar graphs illustrating the distributions of various immune cell types across different samples were generated. The proportions of different immune cell types in the vitiligo and control groups were compared with violin plots. Additionally, heatmaps showing the correlations among 22 types of infiltrating immune cells were generated with the “corrplot” R package.

2.7 Clinical validation of core genes

2.7.1 Sample collection

This study received approval from the Medical Ethics Committee of the Air Force Medical Center of the Chinese People’s Liberation Army (Approval No. 2023-67-PJ01) and followed ethical guidelines consistent with the principles established by the Declaration of Helsinki. All participants provided written informed consent. A total of 30 patients with vitiligo and 30 healthy controls participated in this study. Our research primarily focuses on individuals with the disease who have not undergone systemic treatment. The inclusion criteria are based on the consensus for diagnosing and treating vitiligo (21): (1) skin lesions presented as depigmented white patches of varying sizes and shapes, with normal or increased pigmentation around the edges; (2) skin lesions commonly occurred on the face, neck, hands, torso, and oral and mucosal tissues and the surrounding skin and were frequently found in traumatized areas, with hairs in the white patches typically turning white; (3) other pigmentation disorders or depigmentation diseases were excluded; (4) bright white fluorescence was observed in the white patches under Wood’s lamp examination; (5) disease duration of over 2 years; (6) absence of comorbid autoimmune diseases; and (7) no history of systemic or immunosuppressive therapies.

2.7.2 RNA extraction and quantitative RT–PCR

Total RNA was extracted from peripheral blood using a whole-blood total RNA extraction kit (Simgen, China). The concentration and purity of the extracted RNA were assessed with a NanoDrop

2000 spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA). The assessment focused on A260/A280 ratios that fell within the range of 1.8 to 2.0. The primer sequences of the hub genes used in qRT-PCR are shown in [Table 1](#).

2.8 Skin dataset validation of the hub genes

2.8.1 Merging multiple datasets and removing batch effects

We first merged GSE53146 (GPL14951), GSE75819 (GPL6884) and GSE205155 (GPL14550) using the *inSilicoMerging* package in R (A total of 18 cases of vitiligo patients and 12 healthy controls) ([22](#)), followed by the *Combat* package to remove batch effects ([23](#)).

2.8.2 Expression of the hub genes in the skin-derived datasets

The expression of the hub genes in the skin-derived datasets was calculated via *t* tests.

2.8.3 Receiver operating characteristic curve analysis

The diagnostic performance of the hub genes was assessed via ROC curve analysis.

2.8.4 Single-gene GSEA

We conducted single-gene GSEA to explore potential pathways regulated by the hub genes.

2.9 Statistical analysis

GraphPad Prism (version 9.5.0, San Diego, California, USA) was used to perform two-sided Student's *t* tests to compare the differences between the disease group and the normal group. The *t* value and *P* value for each gene were calculated. A *P* value less than 0.05 was considered to indicate a significant difference between the normal group and the vitiligo group.

3 Results

3.1 Identification of DEGs

In this study, we identified 347 DEGs, including 154 that were upregulated and 193 that were downregulated, from a dataset consisting of 8,621 samples. This dataset included 4 blood samples from healthy individuals and 8 blood samples from patients with vitiligo, sourced from GSE90880. Additionally, 2423 DEGs (398 upregulated genes and 2025 downregulated genes) were selected from a dataset of 20134 samples (including 6 blood samples from healthy individuals and 8 blood samples from vitiligo patients) obtained from GSE80009. A volcano plot of the DEGs was created with thresholds of log2FC greater than 0.5 and an adjusted *p* value

less than 0.05 ([Figures 2A, C](#)). Additionally, a heatmap was generated to illustrate the top 30 upregulated and top 30 downregulated genes ([Figures 2B, D](#)).

3.2 Weighted gene co-expression network analysis

We utilized WGCNA to detect gene modules associated with specific traits. Initially, we calculated a matrix of similarities and transformed it into an adjacency matrix, applying an optimal soft threshold ($\beta = 12$) ([Figures 3A, B](#)). Ten distinct gene modules were identified in the analysis ([Figure 3C](#)). We identified the “brown” module (correlation coefficient = 0.92, *p*-value = 0.06) and the “blue” module (correlation coefficient = 0.91, *p*-value = 0.07) as the modules most clinically relevant to vitiligo on the basis of their associations with the phenotypic traits of the condition ([Figure 3D](#)).

The intersection of the DEGs identified by *limma* in GSE90880 and GSE80009 with the key module genes identified by WGCNA via Venn analysis resulted in the identification of 154 core genes.

To examine the regulatory roles of the 154 core genes in disease development, we compared the DEGs from GSE90880 with those from GSE80009, which were identified via the *limma* package, along with key module genes identified via WGCNA, ultimately leading to the selection of the 154 core genes ([Figure 4A](#)). We subsequently conducted GO and KEGG pathway analyses on the 154 core genes to identify the biological functions of these DEGs. KEGG analysis revealed that the DEGs were associated primarily with pathways such as Epstein–Barr virus infection; the cell cycle; the NF- κ B signalling pathway; parathyroid hormone synthesis, secretion and action; and the spliceosome ([Figures 4B, C](#)). GO analysis revealed enrichment of DEGs in pathways related to negative regulation of cellular processes, the nuclear lumen, negative regulation of metabolic processes, negative regulation of macromolecule metabolic processes, the nucleoplasm, and negative regulation of cellular processes ([Figures 4D, E](#)).

3.3 Machine learning

First, the LASSO regression algorithm was used to identify 16 hub genes ([Figure 5A](#)). Next, the SVM-RFE algorithm was subsequently employed to identify 8 hub genes ([Figure 5B](#)). The RF algorithm was used for identification ([Figure 5C](#)). We visualized the overlapping hub genes from the three algorithms using a Venn diagram, which revealed 2 hub genes: HMGA1 and PSMD13 ([Figure 5D](#)). Finally, we conducted ROC curve analysis on the 2 hub genes, yielding an AUC of 0.92 for HMGA1 and 0.87 for PSMD13 ([Figures 5E, F](#)).

3.4 CIBERSORT

Because the hub genes were found to be enriched in immune-related pathways, we explored the immune landscape in the training

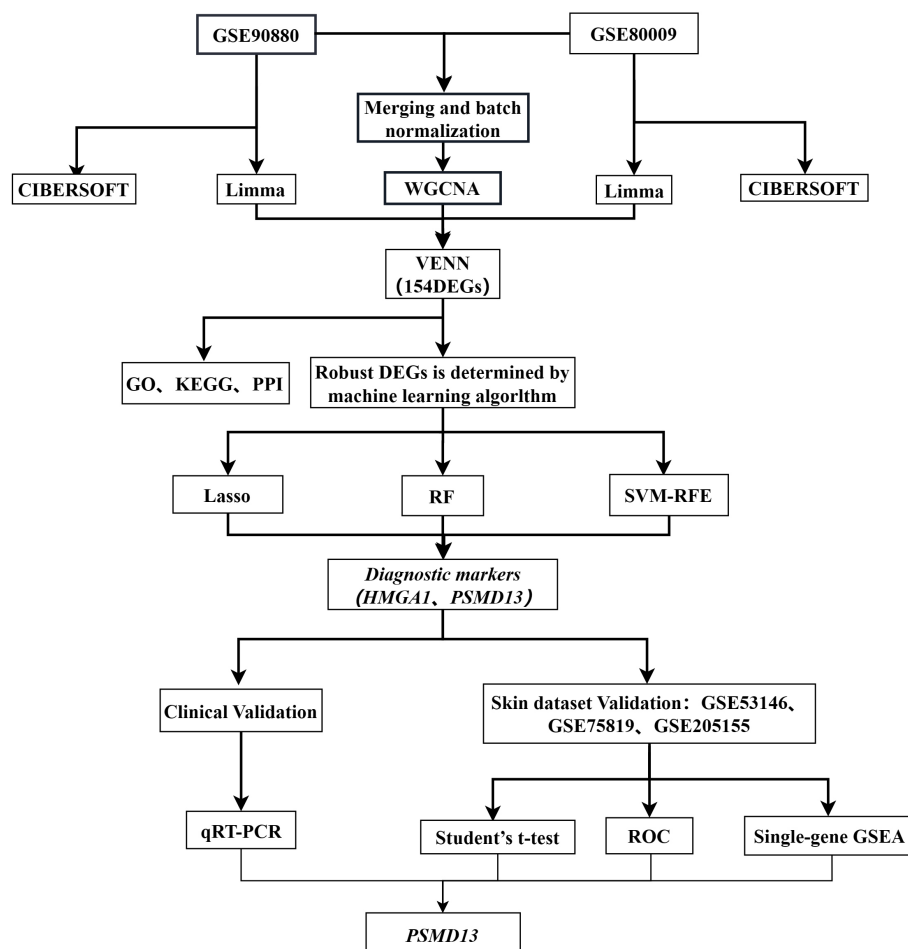


FIGURE 1
Study flowchart.

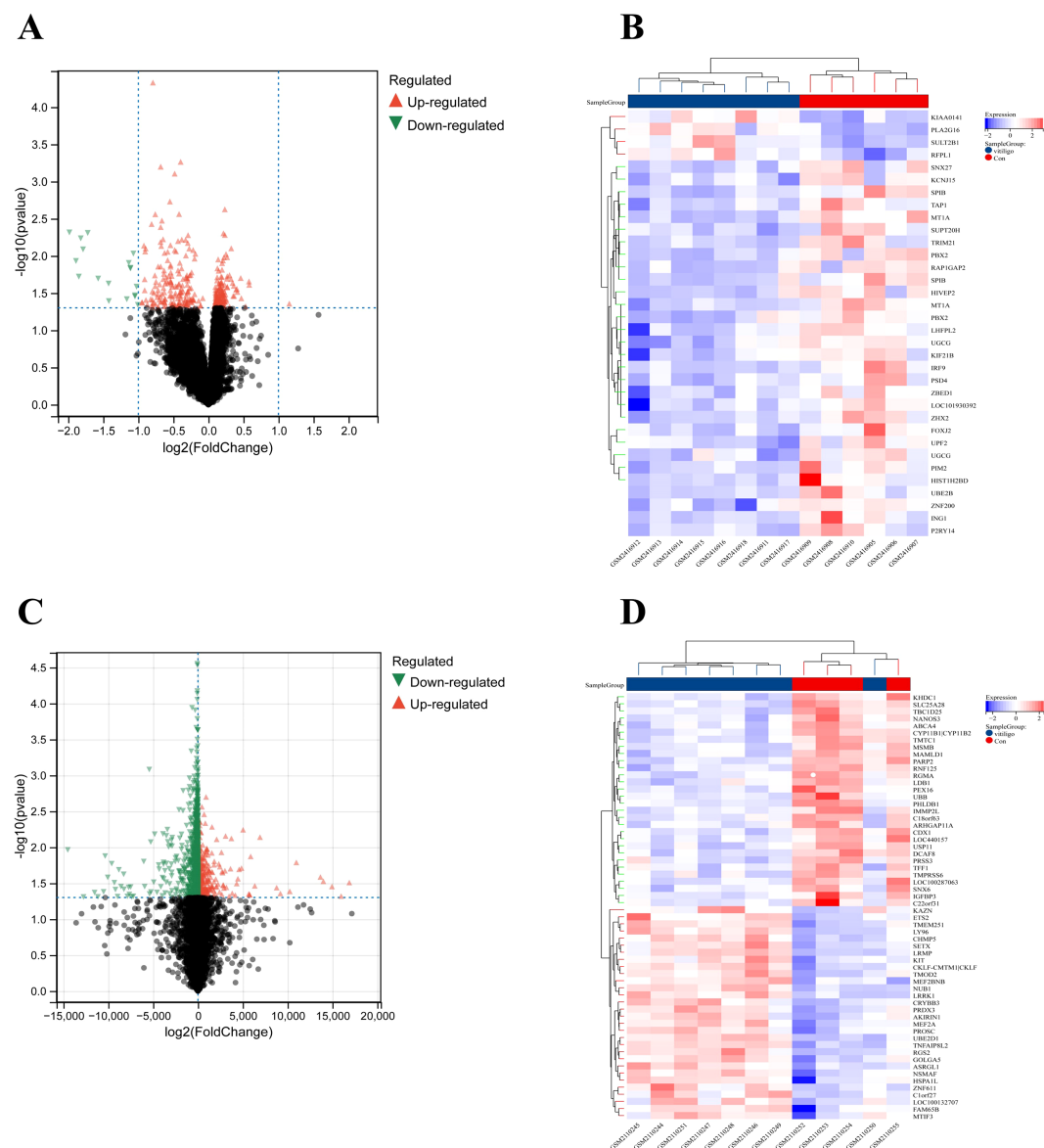
datasets via the CIBERSORT algorithm to better elucidate the role of immune regulation in vitiligo pathogenesis. Figure 6 shows the infiltration levels of 22 immune cell types in both the vitiligo and normal groups. The bar plot visually depicts the proportions of immune cells in each sample from the GSE90880 and GSE80009 datasets (Figures 6A, B). The violin plots displaying differences in immune cell infiltration revealed that, compared with normal control samples, resting natural killer (NK) cells ($p < 0.05$) exhibited greater infiltration in the GSE90880 dataset (Figure 6C). In the GSE80009 dataset, mast cells were activated ($p < 0.01$), and neutrophils ($p < 0.05$) had increased infiltration (Figure 6D).

3.5 qRT-PCR

The expression levels of the overlapping hub genes *HMGA1* and *PSMD13* in the peripheral blood of vitiligo patients and control patients were assessed via qRT-PCR. *HMGA1* and *PSMD13* expression levels were significantly lower in the vitiligo group than in the control group (Figure 7A, $P = 0.015$; Figure 7B, $P = 0.0077$), which is consistent with the results of the bioinformatics analysis.

3.6 Validation of the hub genes in the skin-derived dataset

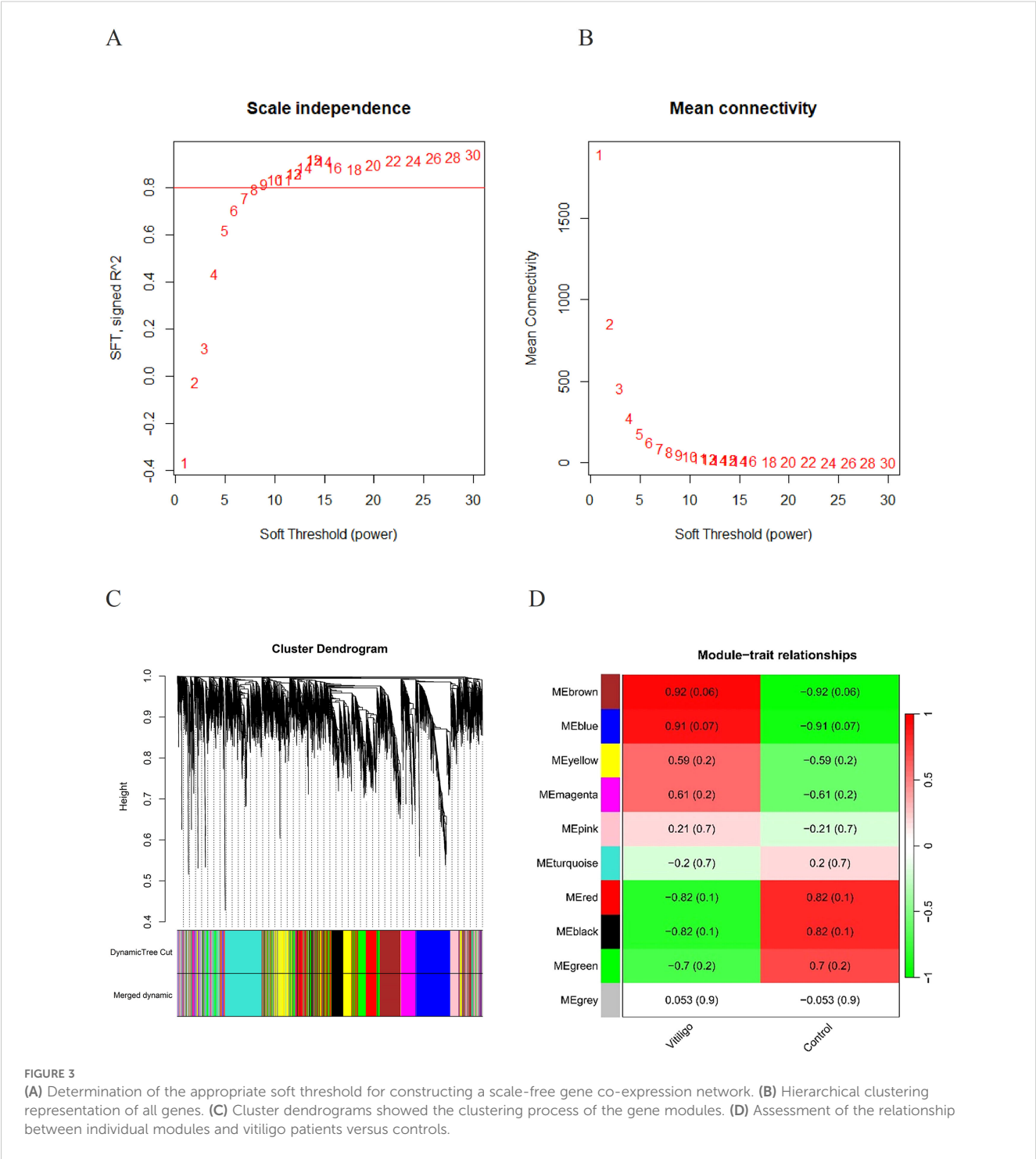
After the batch effect was removed, a boxplot was generated, which suggested that the data distribution among the datasets converged and that the median was on a line (Figure 8A), and the uniform manifold approximation and projection (UMAP) plot demonstrated that the samples from the different datasets clustered closely together, indicating effective removal of the batch effect (Figure 8B). The changes in *HMGA1* expression in the inflamed skin of vitiligo patients compared with that in healthy skin were not significant, but *PSMD13* expression showed a significant downwards trend (Figures 8C, D). We conducted ROC curve analysis to evaluate the diagnostic potential of the two hub genes, with an area under the curve (AUC) value greater than 0.7 considered to indicate a potential diagnostic marker. In the combined dataset, the AUC value of *HMGA1* was 0.50, and that of *PSMD13* was 0.73 (Figures 8E, F). GSEA revealed that the group with low *HMGA1* expression was enriched in epithelial cell signalling pathways associated with *Helicobacter pylori* infection (Figure 8G). The group with low *PSMD13* expression was



genes was confirmed using clinical blood samples and skin datasets. These findings indicate that these genes may significantly contribute to the pathogenesis of vitiligo.

Nucleotide-binding and oligomerization domain (NOD)-like receptors (NLRs) are intracellular proteins with a central role in innate and adaptive immunity (25). NLRs include inflammasome receptors/sensors leading to the maturation of caspase 1, IL-1 β , IL-18, and gasdermin D to drive inflammation and cell death (26). Additionally, several members of the NLR family act as checkpoints in innate immunity by functioning as negative regulators. Various NLRs regulate the balance between cell death, cell survival, autophagy, mitophagy, and even cellular metabolism (27). Many preclinical studies have confirmed that NLRs are closely related to the pathogenesis of vitiligo, indicating that NLRs can regulate

frontiersin.org



melanocyte apoptosis, thereby mediating the onset of vitiligo (28–31). The NF- κ B pathway is crucial for inducing pro-inflammatory cytokines, chemokines, and other inflammatory mediators in different types of innate immune cells. These inflammatory mediators can cause inflammation directly or indirectly promote the differentiation of inflammatory T cells (32). Research indicates that modulating the NF- κ B pathway may alleviate melanocyte damage, offering new insights for the clinical management of vitiligo (33, 34). Additionally, NLRs can respond to external

signals to activate NF- κ B signaling. They negatively regulate NF- κ B signal transduction by interacting with TNF receptor-associated factor 6 (TRAF6), which results in the secretion of inflammatory factors (35, 36).

PSMD13 (26S proteasome non-ATPase regulatory subunit 13), also known as S11 or Rpn9, is a regulatory subunit of the 26S proteasome. *PSMD13* facilitates the ATP-dependent degradation of ubiquitinated proteins, which is crucial for maintaining cellular function through the removal of damaged proteins, and it is also a

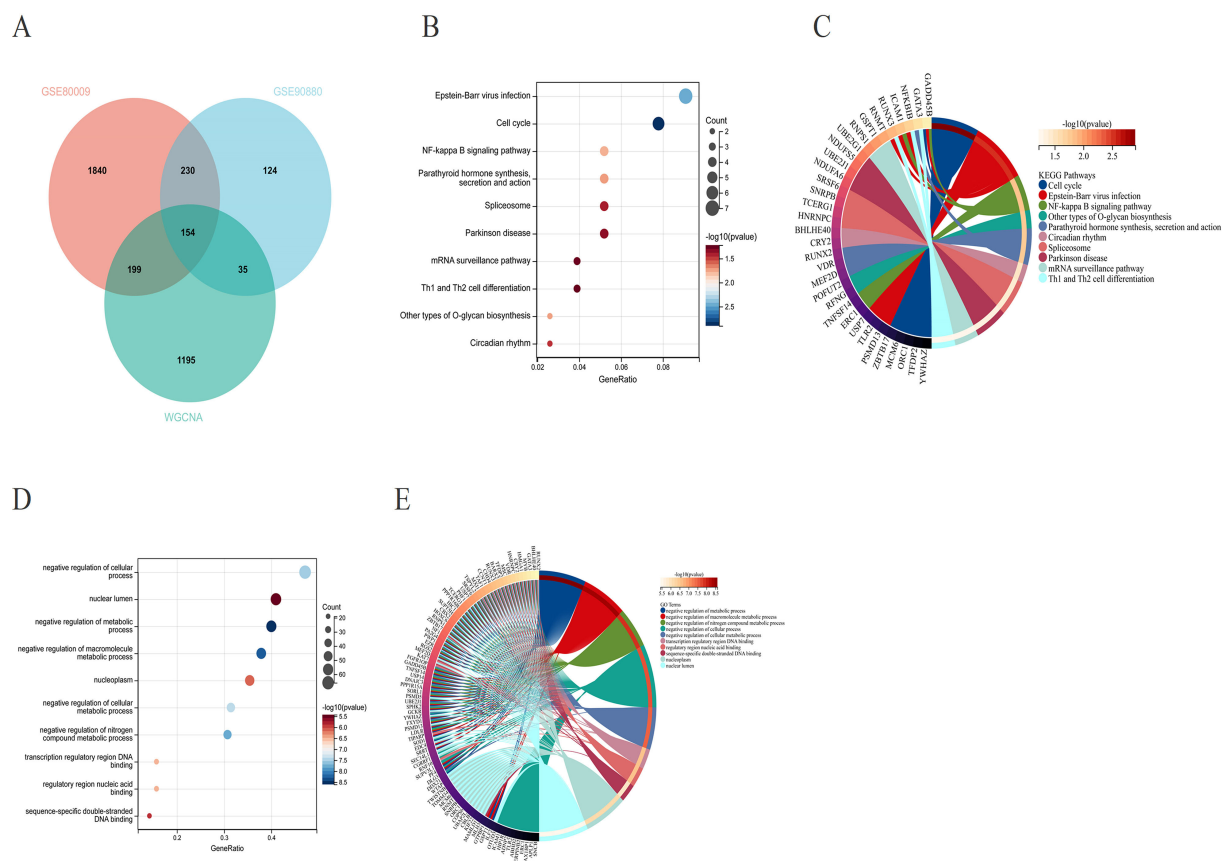


FIGURE 4

(A) Venn diagram illustrating the overlap between the differentially expressed genes from GSE90880 and GSE80009 and the key module genes identified through WGCNA, revealing a total of 154 core genes. (B, C) Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis of the core genes. (D, E) Gene Ontology (GO) analysis of the core genes to assess their functional characteristics.

fundamental complex that helps regulate T-cell function (37). *PSMD13* is involved in processes such as cell cycle regulation and DNA replication, which are essential for maintaining cellular health and longevity (38, 39). One study revealed that *PSMD13* inhibits the NF- κ B pathway, indicating its significant role in innate immune responses (40, 41). In addition, *PSMD13* is involved in the ageing process and is related to the human lifespan (42). A decrease in *PSMD13* expression weakens the resistance of cells to emergency injury and megakaryocyte differentiation (43). Our study revealed a decrease in *PSMD13* expression within the blood samples of vitiligo patients, corroborated by a consistent downwards trend observed in follow-up validations using lesional skin datasets. Further insights gained from single-gene GSEA of patient lesional skin datasets implicate the potential role of the Nod-like receptor signalling pathway in the manifestation of characteristic skin lesions in vitiligo patients. TAK1 plays a crucial role in the NF- κ B signaling pathway. It can induce pyroptosis by facilitating the nuclear translocation of NF- κ B p65 and activating the NOD-like receptor pyrin domain-containing 3 (NLRP3) inflammasome (44, 45). Studies have demonstrated that *PSMD13* inhibits the development of NF- κ B signaling pathway neuroinflammation by targeting TAK1 (40). In summary, *PSMD13* may inhibit NF- κ B signaling through the degradation of TAK1 or kinases associated

with the Nod-like receptor pathway. Additionally, its proteasome function might regulate the stability of NLRs, which could reduce melanocyte damage caused by the inflammasome.

The *HMGA1* protein is a nuclear architectural factor belonging to the non-histone chromosomal binding protein superfamily. *HMGA1* can resist cell apoptosis and enhance senescence, thereby preventing replication (46, 47). Its downregulation can promote myocyte differentiation and skeletal muscle regeneration (48), potentially serving as a downstream target of DNA damage (49). On the one hand, the protein encoded by the *HMGA1* gene regulates gene transcription activity and chromatin structure in the nucleus and plays an important role in cell proliferation, differentiation and growth (50, 51). Our results suggest that in vitiligo, the abnormal expression of *HMGA1* may lead to changes in the function of pigment cells, increasing their vulnerability to attack by immune cells or triggering an autoimmune response, thus exacerbating vitiligo. In addition to the NF- κ B signalling pathway, the Epstein–Barr virus infection pathway is also relevant to the pathogenesis of vitiligo according to KEGG analysis. Although this pathway is associated primarily with viral infection, its role in immune regulation and cell proliferation processes may indirectly influence the immunopathological mechanisms of vitiligo (52). On the other hand, the *PSMD13* gene encodes a protein

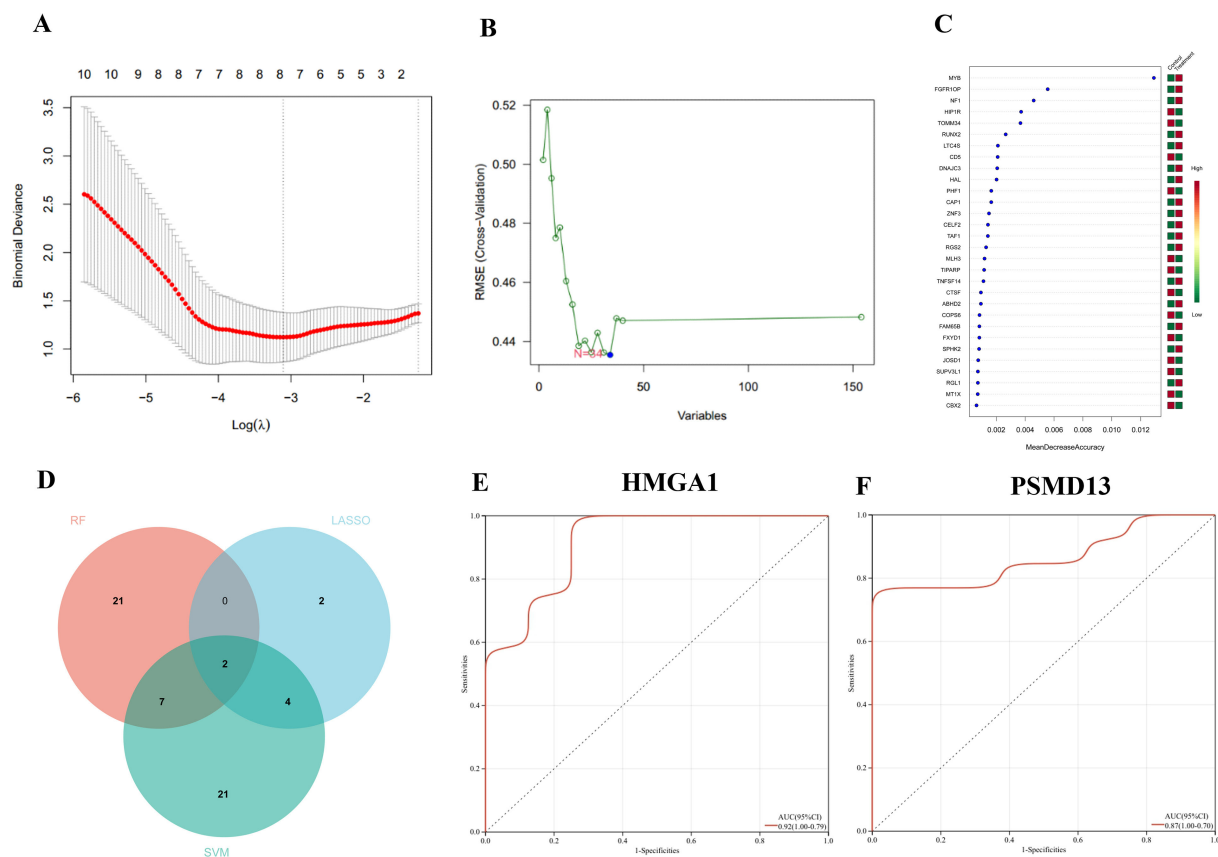


FIGURE 5
Machine learning algorithms: (A) least absolute shrinkage and selection operator (LASSO) regression, (B) support vector machine with recursive feature elimination (SVM-RFE), and (C) the random forest (RF) algorithm. (D) Venn diagram illustrating the overlapping and distinct genes identified by LASSO, SVM-RFE, and RF. (E, F) ROC curve of hub genes *HMGA1* and *PSMD13*.

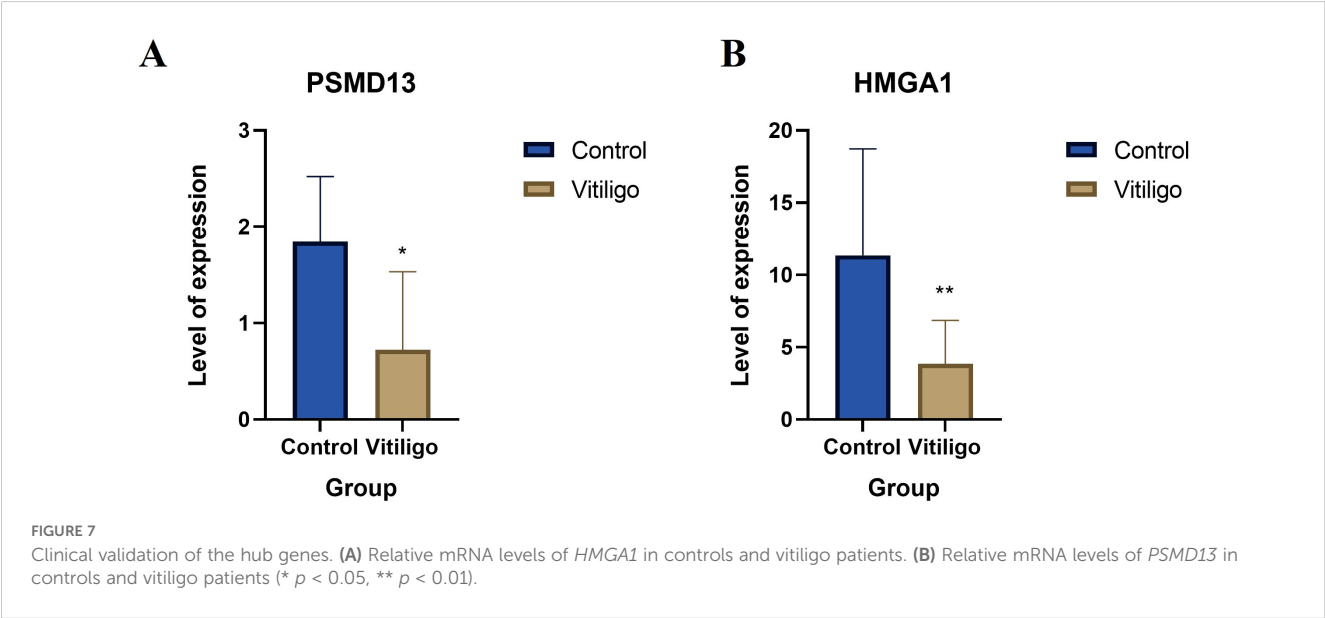
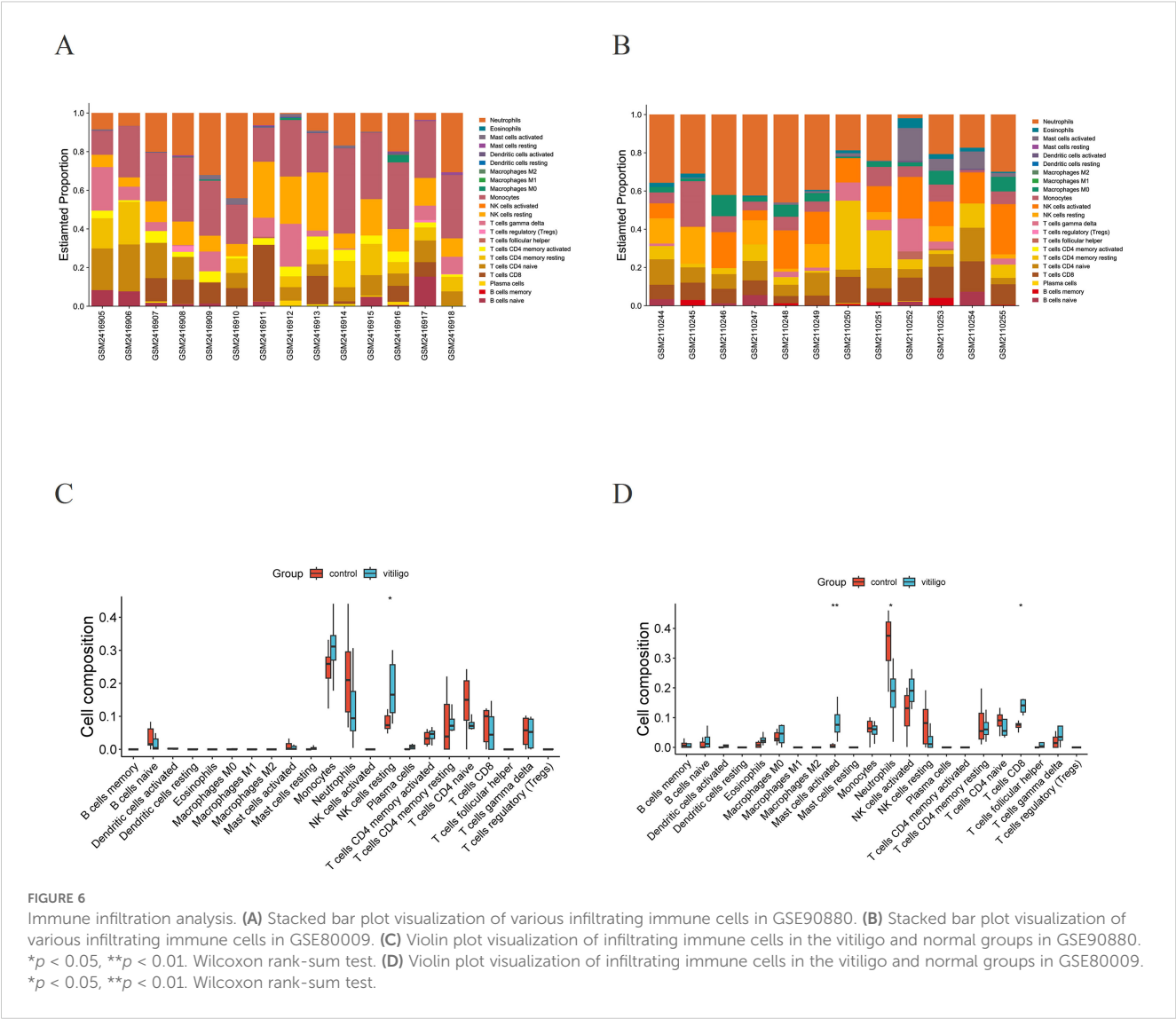
involved in regulating the cell cycle and the NF- κ B signalling pathway, both of which are critical in immune responses and apoptosis (41). The NF- κ B signalling pathway plays a crucial role in regulating inflammation, and its abnormal activation is associated with various autoimmune diseases, including vitiligo. Studies indicate that aberrant activation of the NF- κ B signalling pathway may lead immune cells to attack melanocytes, affecting their normal function and promoting the onset and progression of vitiligo (33, 53).

The role of immune cells in vitiligo cannot be overlooked (54). NK cells, activated mast cells, neutrophils, and other immune cells play crucial roles in the process of immune inflammation. These cells can influence the function and survival of melanocytes directly or indirectly through the release of inflammatory mediators and the activation of immune responses, thereby participating in the pathological processes of vitiligo (55–57).

Our analysis of blood microarray data revealed that reduced expression levels of the *PSMD13* and *HMGA1* genes influence cellular processes such as the cell cycle and NF- κ B signalling, as well as nuclear transcriptional activities. These effects could significantly impact immune-inflammatory responses and the function of pigment cells in individuals with vitiligo. The abnormal

activation of KEGG pathways such as the NF- κ B signalling pathway and the Epstein–Barr virus infection pathway, as well as the activation state of immune cells, are also closely related to the pathogenesis of vitiligo. We verified the expression of the hub genes in blood via qRT–PCR. In addition, *PSMD13* significantly decreased in the lesioned skin of vitiligo patients, suggesting that *PSMD13* is involved in the skin lesion development process through the NOD-like-receptor signalling pathway, but *HMGA1* did not significantly decrease in the lesioned skin. Our study explored the potential of *HMGA1* and *PSMD13* as markers for vitiligo; these results validate our findings and have practical implications for the diagnosis and treatment of vitiligo. Future research should elucidate the specific roles of these molecules and pathways in the development of vitiligo to provide a stronger theoretical foundation for the development of new treatment strategies.

This study identified *PSMD13* and *HMGA1* as potential hub genes linked to vitiligo using integrated bioinformatics and experimental validation. While the exact immune mechanisms of *PSMD13* are unclear, our findings lay a strong foundation for future studies on its functional role, especially in the Nod-like receptor signaling pathway. Furthermore, although the module-trait correlations in WGCNA were only marginally significant ($P =$



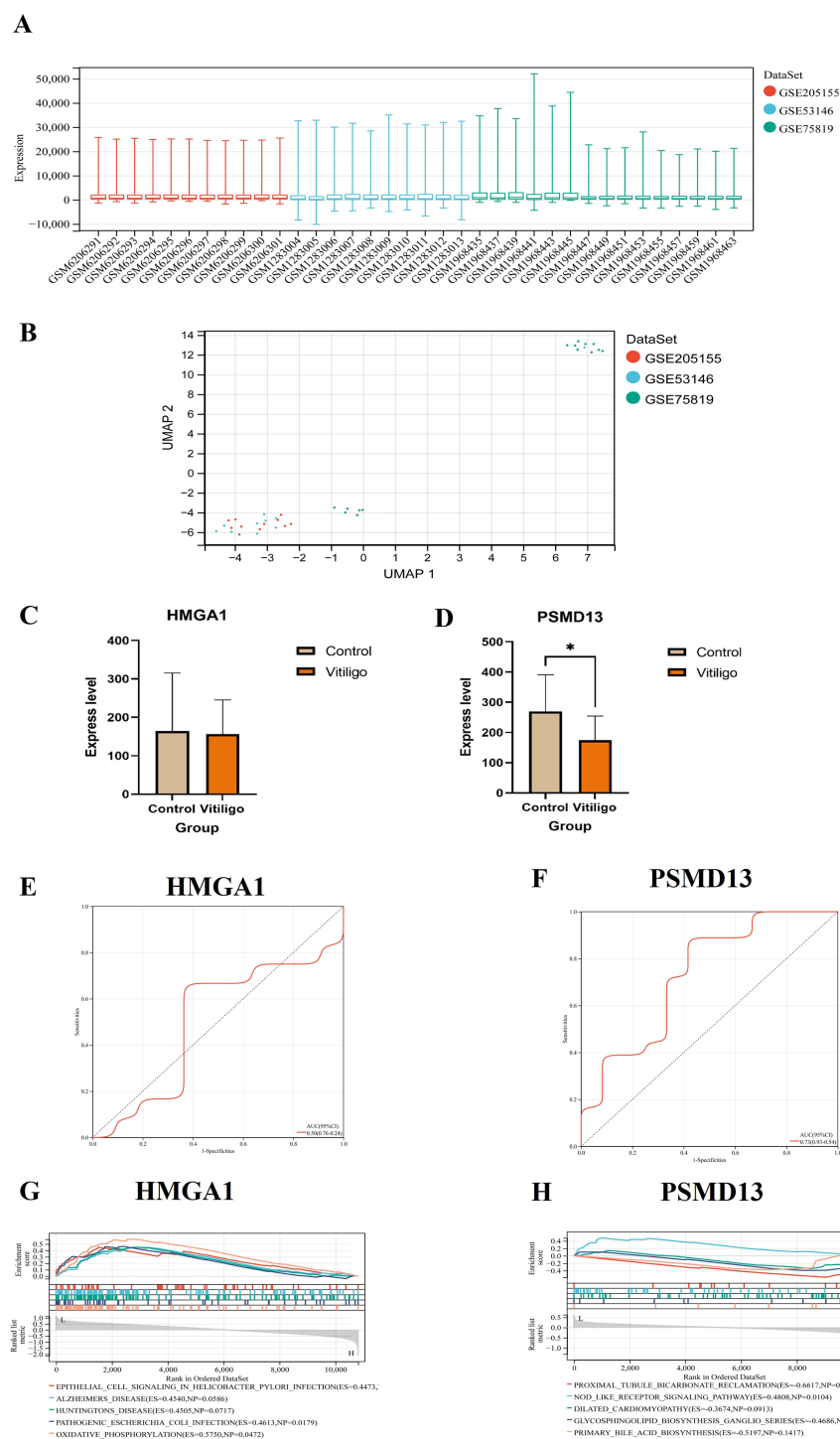


FIGURE 8

Validation of the hub genes in the skin-derived dataset. (A) Boxplot after removal of the batch effect. (B) UMAP plot after removal of the batch effect. (C, D) Expression of hub genes and validation of skin-derived datasets (* $p < 0.05$). (E, F) Hub genes in the skin-derived datasets were analysed via ROC curves. (G, H) GSEA of the hub genes.

0.06 and $P = 0.07$), downstream analyses, such as functional enrichment and immune cell association, strongly supported the relevance of these modules. These limitations underscore the exploratory nature of this study and highlight the necessity for

further validation using larger cohorts and functional assays. Despite these constraints, this study provides valuable insights into the molecular mechanisms underlying vitiligo and suggests novel targets for future diagnostic and therapeutic approaches.

TABLE 1 Sequences of the primers utilized for quantitative real-time PCR.

Gene	Forwards primer sequence	Reverse primer sequence
<i>HMGA1</i>	AGCGAAGTGCCAACACCTAAG	TGGTGGTTTTCCGGGTCTTG
<i>PSMD13</i>	CAGATGACTGATCCTAATGTGGC	CCAGGAAGGTTGTTGAGCATTT
<i>GAPDH</i>	GGCACAGTCAAGGCTGAGAATG	ATGGTGGTGAAGACGCCAGTA

5 Conclusions

In summary, we identified two vitiligo-associated genes, *HMGA1* and *PSMD13*, via WGCNA and machine learning approaches on datasets GSE80009 and GSE90880. We verified this trend by collecting blood samples from vitiligo patients and conducting qRT-PCR analyses of the identified hub genes. Additionally, we gathered microarray datasets (GSE53146 and GSE75819) from affected skin areas of these patients along with a control microarray dataset (GSE205155) from healthy skin for comprehensive expression and GSEA analyses of the hub genes. Our findings indicate that *PSMD13* plays a role in the skin lesion development process via the Nod-like receptor signalling pathway, whereas *HMGA1* does not seem to be directly involved in skin lesions according to our data. Despite their potential importance, research into the specific functions of *HMGA1* and *PSMD13* within the context of vitiligo remains scarce. Our work underscores the importance of *HMGA1* and *PSMD13* in the pathophysiological mechanisms underlying vitiligo, potentially opening new avenues for both diagnostic and treatment strategies for this condition.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: <https://www.ncbi.nlm.nih.gov/geo/>.

Ethics statement

The studies involving humans were approved by Medical Ethics Committee of the Air Force Medical Center of the Chinese People's Liberation Army. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

QW: Methodology, Writing – original draft, Writing – review & editing. JY: Writing – original draft, Conceptualization, Data curation, Investigation, Software. MZ: Writing – original draft,

Supervision, Validation. HJ: Writing – original draft, Formal analysis, Project administration. HL: Formal analysis, Resources, Writing – review & editing. YW: Writing – review & editing, Funding acquisition, Methodology, Writing – original draft.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. Clinical Project at the Air Force Medical University: 2022LC2203.

Acknowledgments

We extend our gratitude to the contributors of the GEO (<https://www.ncbi.nlm.nih.gov/geo/>), STRING (<https://cn.string-db.org/>), GSEA (<https://www.ncbi.nlm.nih.gov/geo/>), and Cibersort (<https://cibersortx.stanford.edu>) databases for sharing their data.

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.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

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.

References

1. Frisoli ML, Essien K, Harris JE. Vitiligo: mechanisms of pathogenesis and treatment. *Annu Rev Immunol.* (2020) 38:621–48. doi: 10.1146/annurev-immunol-100919-023531
2. Manoj R, Singh S, Kothari R, Gupta A. Vitiligo is a condition in which the skin loses its color or pigmentation and becomes pale or white. *J Am Acad Dermatol.* (2024) 90:1106–14. doi: 10.1016/j.jaad.2023.12.040
3. Akl J, Lee S, Ju HJ, Parisi R, Kim JY, Jeon JJ, et al. Estimating the burden of vitiligo: a systematic review and modelling study. *Lancet Public Health.* (2024) 9:e386–96. doi: 10.1016/S2468-2667(24)00026-4
4. Dabas G, Vinay K, Parsad D, Kumar A, Kumaran MS. Psychological disturbances in patients with pigmentary disorders: a cross-sectional study. *J Eur Acad Dermatol Venereol.* (2020) 34:392–9. doi: 10.1111/jdv.15987
5. Lai YC, Yew YW, Kennedy C, Schwartz RA. Vitiligo and depression: a systematic review and meta-analysis of observational studies. *Br J Dermatol.* (2017) 177:708–18. doi: 10.1111/bjd.177.issue-3
6. Li S, Kang P, Zhang W, Jian Z, Zhang Q, Yi X, et al. Activated NLR family pyrin domain containing 3 (NLRP3) inflammasome in keratinocytes promotes cutaneous T-cell response in patients with vitiligo. *J Allergy Clin Immunol.* (2020) 145:632–45. doi: 10.1016/j.jaci.2019.10.036
7. Li S, Zhu G, Yang Y, Jian Z, Guo S, Dai W, et al. Oxidative stress drives CD8(+) T-cell skin trafficking in patients with vitiligo through CXCL16 upregulation by activating the unfolded protein response in keratinocytes. *J Allergy Clin Immunol.* (2017) 140:177–189.e9. doi: 10.1016/j.jaci.2016.10.013
8. Tulic MK, Cavazza E, Cheli Y, Jacquel A, Luci C, Cardot-Leccia N, et al. Innate lymphocyte-induced CXCR3B-mediated melanocyte apoptosis is a potential initiator of T-cell autoreactivity in vitiligo. *Nat Commun.* (2019) 10(1):2178. doi: 10.1038/s41467-019-09963-8
9. Roberts GHL, Paul S, Yorgov D, Santorico SA, Spritz RA. Family clustering of autoimmune vitiligo results principally from polygenic inheritance of common risk alleles. *Am J Hum Genet.* (2019) 105:364–72. doi: 10.1016/j.ajhg.2019.06.013
10. Ezzedine K, Whitton M, Pinart M. Interventions for vitiligo. *Jama.* (2016) 316:1708–9. doi: 10.1001/jama.2016.12399
11. Essien KI, Katz EL, Strassner JP, Harris JE. Regulatory T cells require CCR6 for skin migration and local suppression of vitiligo. *J Invest Dermatol.* (2022) 142:3158–3166.e7. doi: 10.1016/j.jid.2022.05.1090
12. Rosmarin D, Passeron T, Pandya AG, Grimes P, Harris JE, Desai SR, et al. Two phase 3, randomized, controlled trials of ruxolitinib cream for vitiligo. *N Engl J Med.* (2022) 387:1445–55. doi: 10.1056/NEJMoa2118828
13. Rodrigues T, Bernardes GJL. Machine learning for target discovery in drug development. *Curr Opin Chem Biol.* (2020) 56:16–22. doi: 10.1016/j.cbpa.2019.10.003
14. Malik MA, Faraone SV, Michael T, Haavik J. Use of big data and machine learning algorithms to extract possible treatment targets in neurodevelopmental disorders. *Pharmacol Ther.* (2023) 250:108530. doi: 10.1016/j.pharmthera.2023.108530
15. Kakati T, Bhattacharyya DK, Barah P, Kalita JK. Comparison of methods for differential co-expression analysis for disease biomarker prediction. *Comput Biol Med.* (2019) 113:103380. doi: 10.1016/j.compbiomed.2019.103380
16. Feng S, Xu Y, Dai Z, Yin H, Zhang K, Shen Y. Integrative analysis from multicenter studies identifies a WGCNA-derived cancer-associated fibroblast signature for ovarian cancer. *Front Immunol.* (2022) 13:951582. doi: 10.3389/fimmu.2022.951582
17. Zhao X, Zhang L, Wang J, Zhang M, Song Z, Ni B, et al. Identification of key biomarkers and immune infiltration in systemic lupus erythematosus by integrated bioinformatics analysis. *J Transl Med.* (2021) 19:35. doi: 10.1186/s12967-020-02698-x
18. Rezaei Z, Ranjbaran J, Safarpour H, Nomiri S, Salmani F, Chamani E, et al. Identification of early diagnostic biomarkers via WGCNA in gastric cancer. *BioMed Pharmacother.* (2022) 145:112477. doi: 10.1016/j.biopha.2021.112477
19. Komorowski M, et al. Sepsis biomarkers and diagnostic tools with a focus on machine learning. *EBioMedicine.* (2022) 86:104394. doi: 10.1016/j.ebiom.2022.104394
20. Li R. Data mining and machine learning methods for dementia research. *Methods Mol Biol.* (2018) 1750:363–70. doi: 10.1007/978-1-4939-7704-8_25
21. Böhm M, Schunter JA, Fritz K, Salavastru C, Dargatz S, Augustin M, et al. S1 Guideline: Diagnosis and therapy of vitiligo. *J Dtsch Dermatol Ges.* (2022) 20:365–78. doi: 10.1111/ddg.14713
22. Taminiau J, Meganck S, Lazar C, Steenhoff D, Coletta A, Molter C, et al. Unlocking the potential of publicly available microarray data using inSilicoDb and inSilicoMerging R/Bioconductor packages. *BMC Bioinf.* (2012) 13:335. doi: 10.1186/1471-2105-13-335
23. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics.* (2007) 8:118–27. doi: 10.1093/biostatistics/kxj037
24. Chen J, Li S, Li C. Mechanisms of melanocyte death in vitiligo. *Med Res Rev.* (2021) 41:1138–66. doi: 10.1002/med.21754
25. Zhou Y, Yu S, Zhang W. NOD-like receptor signaling pathway in gastrointestinal inflammatory diseases and cancers. *Int J Mol Sci.* (2023) 24(19):14511. doi: 10.3390/ijms241914511
26. Chou WC, Jha S, Linhoff MW, Ting JP. The NLR gene family: from discovery to present day. *Nat Rev Immunol.* (2023) 23:635–54. doi: 10.1038/s41577-023-00849-x
27. Truax AD, Chen L, Tam JW, Cheng N, Guo H, Koblansky AA, et al. The inhibitory innate immune sensor NLRP12 maintains a threshold against obesity by regulating gut microbiota homeostasis. *Cell Host Microbe.* (2018) 24:364–378.e6. doi: 10.1016/j.chom.2018.08.009
28. Rajendiran KS, Rajappa M, Chandrashekar L, Thappa DM. Association of Nod-like receptor protein-1 (rs2670660) and Toll-like receptor-4 (rs4986790) with non-segmental vitiligo: A case-control study in South Indian population. *Int J Immunogenet.* (2019) 46:321–30. doi: 10.1111/iji.12429
29. Silverberg NB. Update on childhood vitiligo. *Curr Opin Pediatr.* (2010) 22:445–52. doi: 10.1097/MOP.0b013e32833b6ac3
30. Marie J, Kovacs D, Pain C, Jouary T, Cota C, Vergier B, et al. Inflammasome activation and vitiligo/nonsegmental vitiligo progression. *Br J Dermatol.* (2014) 170:816–23. doi: 10.1111/bjd.2014.170.issue-4
31. Calabrese L, Fiocco Z, Mellett M, Aoki R, Rubegni P, French LE, et al. Role of the NLRP1 inflammasome in skin cancer and inflammatory skin diseases. *Br J Dermatol.* (2024) 190:305–15. doi: 10.1093/bjd/ljad421
32. Lawrence T. The nuclear factor NF-kappaB pathway in inflammation. *Cold Spring Harb Perspect Biol.* (2009) 1(6):a001651. doi: 10.1101/cshperspect.a001651
33. Wang J, Lin F, Zhang W, Xu A, DeGiorgis J, Lu H, et al. Novel approaches to vitiligo treatment via modulation of mTOR and NF-kB pathways in human skin melanocytes. *Int J Biol Sci.* (2017) 13:391–400. doi: 10.7150/ijbs.17318
34. Zhou J, Ling J, Song J, Wang Y, Feng B, Ping F. Interleukin 10 protects primary melanocyte by activation of Stat-3 and PI3K/Akt/NF-kB signaling pathways. *Cytokine.* (2016) 83:275–81. doi: 10.1016/j.cyt.2016.05.013
35. Ogura Y, Inohara N, Benito A, Chen FF, Yamaoka S, Nunez G. Nod2, a Nod1/ Apaf-1 family member that is restricted to monocytes and activates NF-kappaB. *J Biol Chem.* (2001) 276:4812–8. doi: 10.1074/jbc.M008072200
36. Morrison HA, Trusiano B, Rowe AJ, Allen IC. Negative regulatory NLRs mitigate inflammation via NF-kB pathway signaling in inflammatory bowel disease. *BioMed J.* (2023) 46:100616. doi: 10.1016/j.bj.2023.100616
37. Minelli A, Magri C, Barbon A, Bonvicini C, Segala M, Congiu C, et al. Proteasome system dysregulation and treatment resistance mechanisms in major depressive disorder. *Transl Psychiatry.* (2015) 5:e687. doi: 10.1038/tp.2015.180
38. Williams AJ, Dave JR, Tortella FC. Neuroprotection with the proteasome inhibitor MLN519 in focal ischemic brain injury: relation to nuclear factor kappaB (NF-kappaB), inflammatory gene expression, and leukocyte infiltration. *Neurochem Int.* (2006) 49:106–12. doi: 10.1016/j.neuint.2006.03.018
39. Farout L, Friguet B. Proteasome function in aging and oxidative stress: implications in protein maintenance failure. *Antioxid Redox Signal.* (2006) 8:205–16. doi: 10.1089/ars.2006.8.205
40. Li Q, Chen Y, Wang PF, Sun YA, Xu TJ. PSMD13 inhibits NF-kB pathway by targeting TAK1 for K63-linked ubiquitination in miiuy croaker (*Miichthys miiuy*). *Fish Shellfish Immunol.* (2023) 138:108857. doi: 10.1016/j.fsi.2023.108857
41. Bi W, Zhu LH, Zeng ZF, Jing XN, Liang YR, Guo L, et al. Investigations into the role of 26S proteasome non-ATPase regulatory subunit 13 in neuroinflammation. *Neuroimmunomodulation.* (2014) 21:331–7. doi: 10.1159/000357811
42. Bellizzi D, Dato S, Cavalcante P, Covello G, Di Cianni F, Passarino G, et al. Characterization of a bidirectional promoter shared between two human genes related to aging: SIRT3 and PSMD13. *Genomics.* (2007) 89:143–50. doi: 10.1016/j.ygeno.2006.09.004
43. Gao A, Xu S, Li Q, Zhu C, Wang F, Wang Y, et al. Interleukin-4 weakens resistance to stress injury and megakaryocytic differentiation of hematopoietic stem cells by inhibiting Psmd13 expression. *Sci Rep.* (2023) 13:14253. doi: 10.1038/s41598-023-41479-6
44. Sun W, Wu G, Tian X, Qi C, Liu J, Tong Y, et al. Small molecule activators of TAK1 promotes its activity-dependent ubiquitination and TRAIL-mediated tumor cell death. *Proc Natl Acad Sci U S A.* (2023) 120:e2308079120. doi: 10.1073/pnas.2308079120
45. Xu P, Tao C, Zhu Y, Wang G, Kong L, Li W, et al. TAK1 mediates neuronal pyroptosis in early brain injury after subarachnoid hemorrhage. *J Neuroinflammation.* (2021) 18:188. doi: 10.1186/s12974-021-02226-8
46. Wang L, Zhang J, Xia M, Liu C, Zu X, Zhong J. High mobility group A1 (HMGAI1): structure, biological function, and therapeutic potential. *Int J Biol Sci.* (2022) 18:4414–31. doi: 10.7150/ijbs.72952
47. Reeves R. Molecular biology of HMGA proteins: hubs of nuclear function. *Gene.* (2001) 277:63–81. doi: 10.1016/S0378-1119(01)00689-8
48. Ai J, Zhao F, Zhou X. HMGA1 aggravates oxidative stress injury and inflammatory responses in IL-1 β -induced primary chondrocytes through the JMD3/ZEB1 axis. *Int Arch Allergy Immunol.* (2023) 184:279–90. doi: 10.1159/000526680

49. Pentimalli F, Palmieri D, Pacelli R, Garbi C, Cesari R, Martin E, et al. HMGAI protein is a novel target of the ATM kinase. *Eur J Cancer*. (2008) 44:2668–79. doi: 10.1016/j.ejca.2008.07.033
50. Reeves R, Beckerbauer L. HMG/Y proteins: flexible regulators of transcription and chromatin structure. *Biochim Biophys Acta*. (2001) 1519:13–29. doi: 10.1016/S0167-4781(01)00215-9
51. Diana F, Sgarra R, Manfioletti G, Rustighi A, Poletto D, Sciortino MT, et al. A link between apoptosis and degree of phosphorylation of high mobility group A1a protein in leukemic cells. *J Biol Chem*. (2001) 276:11354–61. doi: 10.1074/jbc.M009521200
52. Zhuang T, Yi X, Chen J, Kang P, Chen X, Chen J, et al. Intracellular virus sensor MDA5 exacerbates vitiligo by inducing the secretion of chemokines in keratinocytes under virus invasion. *Cell Death Dis*. (2020) 11:453. doi: 10.1038/s41419-020-2665-z
53. Kim NH, Jeon S, Lee HJ, Lee AY. Impaired PI3K/Akt activation-mediated NF-kappaB inactivation under elevated TNF-alpha is more vulnerable to apoptosis in vitiliginous keratinocytes. *J Invest Dermatol*. (2007) 127:2612–7. doi: 10.1038/sj.jid.5700900
54. Koike S, Yamasaki K. Melanogenesis connection with innate immunity and toll-like receptors. *Int J Mol Sci*. (2020) 21(24):9769. doi: 10.3390/ijms21249769
55. Chen Y, Griffiths CEM, Bulfone-Paus S. Exploring mast cell-CD8 T cell interactions in inflammatory skin diseases. *Int J Mol Sci*. (2023) 24(2):1564. doi: 10.3390/ijms24021564
56. Solak B, Dikicier BS, Cosansu NC, Erdem T. Neutrophil to lymphocyte ratio in patients with vitiligo. *Postepy Dermatologii I Alergologii*. (2017) 34:468–70. doi: 10.5114/ada.2017.71114
57. Basak PY, Adiloglu AK, Koc IG, Tas T, Akkaya VB. Evaluation of activatory and inhibitory natural killer cell receptors in non-segmental vitiligo: a flow cytometric study. *J Eur Acad Dermatol Venereology*. (2008) 22:970–6. doi: 10.1111/j.1468-3083.2008.02681.x