

## Research Article

# Developing a Diagnostic Model to Predict the Risk of Asthma Based on Ten Macrophage-Related Gene Signatures

Xiaoshun Ai <sup>1</sup>, Hong Shen <sup>2</sup>, Yangyanqiu Wang <sup>3</sup>, Jing Zhuang <sup>4</sup>, Yani Zhou <sup>3</sup>,  
Furong Niu <sup>2</sup> and Qing Zhou <sup>4</sup>

<sup>1</sup>Huzhou First Hospital, Zhebei Mingzhou Hospital, No. 225, Gongyuan Road, Wuxing District, Huzhou Zhejiang Province, China 313000

<sup>2</sup>School of Medicine, Huzhou University, No. 759 Erhuan East Road, Huzhou, Zhejiang Province, China 313000

<sup>3</sup>Huzhou Hospital of Zhejiang University, Affiliated Central Hospital Huzhou University, No.1558, Sanhuan North Road, Wuxing District, Huzhou, Zhejiang Province, China 313000

<sup>4</sup>Huzhou Central Hospital, Affiliated Central Hospital Huzhou University, Key Laboratory of Multiomics Research and Clinical Transformation of Digestive Cancer of Huzhou, No. 1558, Sanhuan North Road, Wuxing District, Huzhou Zhejiang Province, China 313000

Correspondence should be addressed to Furong Niu; niufurong2021@163.com and Qing Zhou; zhouqing20182018@163.com

Received 11 May 2022; Revised 9 October 2022; Accepted 31 October 2022; Published 23 November 2022

Academic Editor: Si Qin

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

**Objective.** Asthma (AS) is a chronic inflammatory disease of the airway, and macrophages contribute to AS remodeling. Our study aims at screening macrophage-related gene signatures to build a risk prediction model and explore its predictive abilities in AS diagnosis. **Methods.** Three microarray datasets were downloaded from the GEO database. The Limma package was used to screen differentially expressed genes (DEGs) between AS and controls. The ssGSEA algorithm was used to determine immune cell proportions. The Pearson correlation coefficient was computed to select the macrophage-related DEGs. The LASSO and RFE algorithms were implemented to filter the macrophage-related DEG signatures to establish a risk prediction model. Receiver operating characteristic (ROC) curves were used to assess the diagnostic ability of the prediction model. Finally, the qPCR was used to detect the expression of selected differential genes in sputum from healthy people and asthmatic patients. **Results.** We obtained 1,189 DEGs between AS and controls from the combined datasets. By evaluating immune cell proportions, macrophages showed a significant difference between the two groups, and 439 DEGs were found to be associated with macrophages. These genes were mainly enriched in the gene ontology-biological process of immune and inflammatory responses, as well as in the KEGG pathways of cytokine-cytokine receptor interaction and biosynthesis of antibiotics. Finally, 10 macrophage-related DEG signatures (*EARS2*, *ATP2A2*, *COLGALT1*, *GART*, *WNT5A*, *AK5*, *ZBTB16*, *CCL17*, *ADORA3*, and *CXCR4*) were screened as an optimized gene set to predict AS diagnosis, and they showed diagnostic abilities with AUCs of 0.968 and 0.875 in ROC curves of combined and validation datasets, respectively. The mRNA expressions of *EARS2*, *ATP2A2*, *COLGALT1*, and *GART* in the control group were higher than in AS group, while the expressions of *WNT5A*, *AK5*, *ZBTB16*, *CCL17*, *ADORA3*, and *CXCR4* in the control group were lower than that in the AS group. **Conclusion.** We proposed a diagnostic model based on 10 macrophage-related genes to predict AS risk.

## 1. Introduction

Asthma (AS) is a respiratory disease that clinically manifests as airway hyperresponsiveness, inflammation, and mucus secretion [1]. As a heterogeneous clinical syndrome, AS affects more than 300 million people worldwide, with an

annually increasing prevalence rate [2]. Environmental factors, comorbidities, genomic factors, and other social determinants are considered to play a synergistic role in the etiology of asthma [3]. Studies on the population genetics of AS have shown that genetic factors contribute to AS pathogenesis, with estimates of heritability ranging from 35 to

95% [4]. Therefore, it is necessary to study the candidate genes related to AS and AS phenotypes as well as their internal molecular mechanisms. Airway and alveolar epithelial cells are the first line of defense of the lung's immune system, protecting against invading pathogens and environmental pollutants [5]. AS is characterized by airway inflammation, and immune cell infiltration and goblet cell proliferation are often observed in the airways of AS patients [6]. The relationship between inflammatory response, immune activation, and AS exacerbation is well established; thus, studies on biomarkers associated with immune response in AS may help to improve clinical outcomes of AS by reducing early inflammation.

Macrophages contribute the largest proportion of leukocytes (accounting for approximately 70% of immune cells) found in alveoli, distal airspaces, and conducting airways [7, 8]. Macrophages are variously involved in AS inflammation, including altering the production of anti-inflammatory cytokines or chemokines and inducing inflammasomes to regulate cellular processes [9]. Although macrophages are abundant in lung tissues, their contribution to AS pathology comes more from functional changes. Studies have reported that macrophage function depends on the polarization state of Th1 and Th2 [10]. M1 macrophages are induced by IFN- $\gamma$  and lipopolysaccharide and function in driving inflammation in response to intracellular pathogens, whereas M2 macrophages are involved in anti-inflammation induced by IL-4 and IL-13 [11]. These findings support the role of macrophages as disease modifiers, biomarkers, and therapeutic targets in AS. However, studies are still needed to explore how cellular signaling and gene signature expression influence the functional response of macrophages.

Therefore, the purpose of this study was to screen differentially expressed genes (DEGs) related to macrophages in AS through samples sourced from a public database. Then, macrophage-related DEG signatures were filtered as an optimized gene set to build a risk prediction model. Finally, the expression of candidate genes and the prediction model's efficiency in AS diagnosis and recognition were validated. The analytical flowchart of this study is summarized in Supplemental Figure 1. This study proposed novel biomarkers associated with macrophages in AS and provides new insights into AS therapeutic strategies by highlighting the potential function of macrophages.

## 2. Methods

**2.1. Data Acquisition.** The expression data were obtained from three microarray datasets in the Gene Expression Omnibus (GEO) (<https://www.ncbi.nlm.nih.gov/>) [12], including GSE137268, GSE148004, and GSE112260. In detail, 54 AS and 15 normal samples detected by the Illumina humanRef-8 v2.0 expression beadchip were collected from GSE63067; 9 AS and 9 normal samples detected by the Agilent-026652 Whole Human Genome Microarray 4  $\times$  44K v2 were obtained from GSE148004 [13], and 4 AS samples and 4 healthy controls were obtained from GSE112260 [14] based on the Affymetrix Human Gene 2.1 ST Array. Among them, GSE137268 and GSE148004 were utilized for analysis, while GSE112260 was used for validation.

**2.2. Screening for DEGs of AS.** Principal component analysis (PCA) was performed to remove the batch effect of samples in GSE137268 and GSE148004 using the R3.6.1 sva package version 3.38.0 [15] (<http://www.bioconductor.org/packages/release/bioc/html/sva.html>). After gathering the combined expression profile data, the R3.6.1 Limma package version 3.34.7 [16] (<https://bioconductor.org/packages/release/bioc/html/limma.html>) was used to screen the DEGs of AS with a standard of fold discovery rate (FDR) < 0.05, and  $|\log_2$  fold change (FC)| > 0.263.

**2.3. Screening of Macrophage-Related DEGs.** To evaluate immune cell infiltration in the combined samples, immunologic signature gene sets were downloaded from the Gene Set Enrichment Analysis website (GSEA, <http://software.broadinstitute.org/gsea/index.jsp>). After this, single-sample gene set enrichment analysis (ssGSEA) [17] was implemented using the gene set variation analysis (GSVA) package version 1.36.3 [18] (<http://www.bioconductor.org/packages/release/bioc/html/GSVA.html>) in R3.6.1, to compare the differences in the proportion of individual immune cells between AS and normal samples.

The correlation between the DEGs of AS and the proportion of macrophages was assessed using the R3.6.1 cor function (<http://77.66.12.57/R-help/cor.test.html>). By calculating the Pearson correlation coefficient (PCC), DEGs with  $P < 0.05$  were determined to be significantly associated with macrophages.

The R package clusterProfiler (<http://bioconductor.org/packages/release/bioc/html/clusterProfiler.html>) was used for analysis of gene ontology biological process (GO-BP) function and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment of macrophage-related DEGs. The  $P$  adjust less than 0.05 and the count value was greater than 1 were considered as the threshold screening criterion.

**2.4. Construction of Protein-Protein Interaction (PPI) Network.** STRING version 11.0 [19] (<http://string-db.org/>) was used to establish the interactions of the coding proteins of macrophage-related DEGs with a combined score threshold of 0.6. A PPI network was constructed and visualized using Cytoscape version 3.6.1 [20] (<http://www.cytoscape.org/>). KEGG pathway enrichment analysis of hub genes in the network was then performed using the R package clusterProfiler.

**2.5. Screening and Verification of Macrophage-Related Gene Signatures.** Two different algorithms were used to screen the DEG signatures from hub genes in the PPI network. Specifically, the lars package version 1.2 [21] (<https://cran.r-project.org/web/packages/lars/index.html>) was used for least absolute shrinkage and selection operator (LASSO) regression analysis on hub genes, and the caret package version 6.0-76 [22] (<https://cran.r-project.org/web/packages/caret>) was used for recursive feature elimination (RFE) to select candidate genes. The intersection DEGs were then determined to be macrophage-related DEG signatures.

The expression data of macrophage-related DEG signatures were extracted from the GSE112260 dataset and compared between AS and the controls. The support vector

machine (SVM) approach [23] was utilized to construct a disease diagnostic classifier (Core: Sigmoid Kernel; Cross: 100-fold cross-validation) using R3.6.1 e1071 version 1.6-8 (<https://cran.r-project.org/web/packages/e1071>). The sensitivity and specificity of the receiver operating characteristic (ROC) curve calculated using R 3.6.1 pROC version 1.12.1 [24] (<https://cran.r-project.org/web/packages/pROC/index.html>) were used to evaluate the performance of the diagnostic model in the combined dataset and validation dataset.

The patients with AS admitted to the Huzhou Traditional Chinese Medicine Hospital and healthy volunteers from the physical examination center in Huzhou Traditional Chinese Medicine Hospital were recruited in the study as subjects. The Ethics Committee of Huzhou Traditional Chinese Medicine Hospital approved this study, and all subjects provided informed consent (N0.2021-030-A) for approval. These are the inclusion criteria: subjects were diagnosed with AS according to National Asthma Education and Prevention Program Coordinating Committee Expert Panel Working Group (NAEPP). The exclusion criteria are as follows: subjects with other respiratory diseases, such as allergic rhinitis, endotracheal disease, bronchial lung cancer, etc. Patients with other malignant tumors or serious cardiopulmonary diseases. The patients had taken bronchodilators, glucocorticoids, and other asthma medications a week earlier. Sputum from asthmatic patients was obtained as follows. After inhalation of hypertonic saline atomization for 15 min, the lungs were tapped, and sputum was extracted and collected in a sterile environment. The clinical characteristics of the patients are shown in Supplemental Table 1.

**2.6. qPCR Analysis.** TRIzol (Invitrogen; Thermo Fisher Scientific, Inc) was added to the cells after washing with PBS. RNA was extracted according to a previously described method. The quality and concentration of RNA were measured by an Infinite M100 PRO (Tecan Group Ltd., China). cDNA was obtained using RRO47A (TAKARA BIO INC, Japan) according to the protocols. Subsequently, real-time qPCR (Funglyn Biotech, Inc, Ontario, Canada) was performed at 50°C for 3 min, followed by 95°C for 3 min, 95°C for 10 s, and 60°C for 30 s; this was repeated for 40 cycles. Information on primers is shown in Supplemental Table 2. The expression levels of miRNAs and genes were measured using the  $2^{-\Delta\Delta C_t}$  method.

### 3. Results

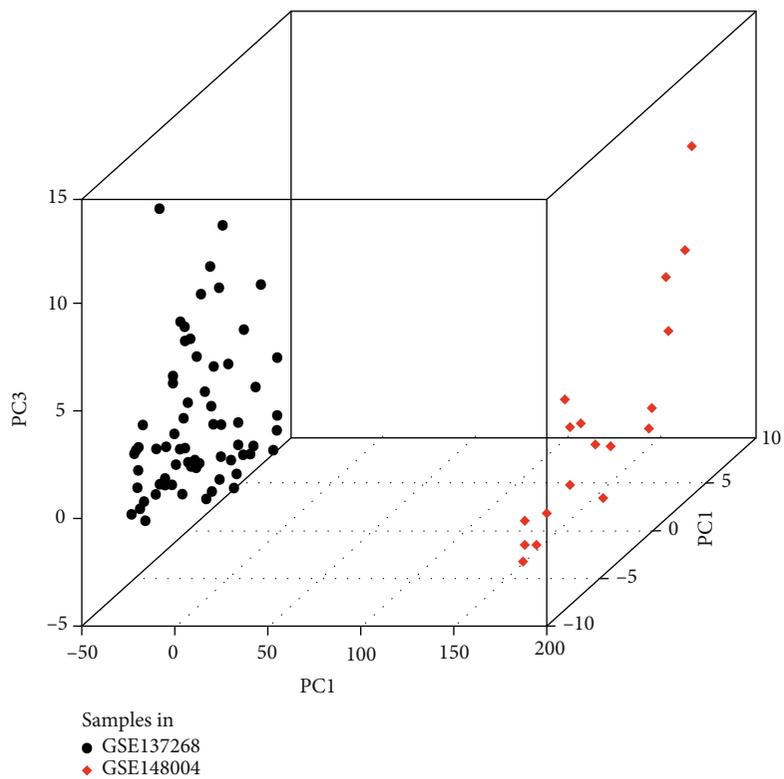
**3.1. Screening of DEGs between AS Samples and Healthy Controls.** We combined GSE137268 and GSE148004 into one dataset and removed the batch effect using the SVA algorithm. The expression levels before and after removal of the batch effect are shown in Supplemental Figure 2. A PCA of the combined samples was further performed (Figures 1(a) and 1(b)), and the results indicated that the combined samples from the two different detection platforms were indistinguishable after batch effect removal and could be applied for further analysis. The Limma package was then used to screen DEGs between AS samples and healthy controls from the combined dataset, and we obtained 1,189

DEGs in total with the corresponding thresholds, as shown in Figure 1(c). The heatmap (Figure 1(d)) showed that DEG expression was significantly different between the AS and control groups, indicating that the screened DEGs had expression features in each group.

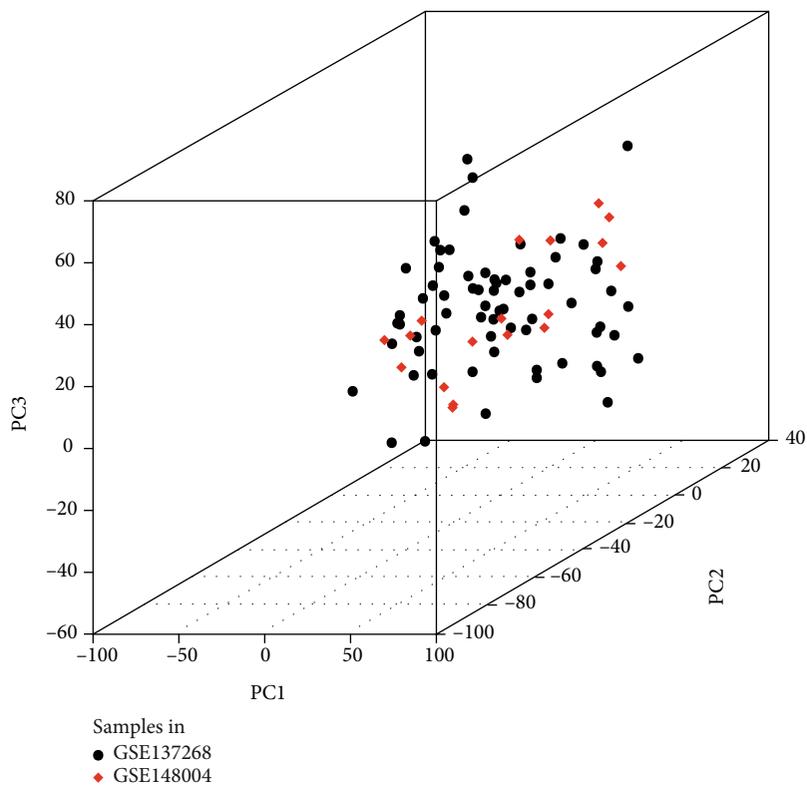
**3.2. Selection of Macrophage-Related DEGs and Functional Enrichment Analysis.** Based on the expression profiling of the combined samples, we obtained the proportions of 28 types of immune cells using the ssGSEA algorithm. By comparing the differences in the proportion of immune cells between AS samples and healthy controls, we found that seven immune cell types were significantly different between the two groups, including macrophages, with a *P* value of 0.011 (Supplemental Figure 3 and Supplemental Table 3). By calculating the PCC between DEGs and cell proportion of macrophages in samples using the *cor* function in R3.6.1, 439 macrophage-related DEGs were screened at a *P* value < 0.05. We then carried out function and pathway enrichment analyses on macrophage-related DEGs using the R package clusterProfiler and filtered 20 biological processes (Figure 2(a)), including regulation of inflammatory response as well as 20 KEGG pathways (Figure 2(b)) including Metabolic pathways, pathways in cancer and cytokine-cytokine receptor interaction, etc., with statistical significance.

**3.3. Establishment and Analysis of PPI Network.** The STRING database was applied to establish the interactions between the coding proteins of macrophage-related DEGs, and a total of 515 pairs of interactions were obtained with combined scores over 0.6. A PPI network comprising 220 gene nodes was constructed, as shown in Figure 3(a). We also analyzed the topology properties of nodes in the network and listed the detailed information of nodes with degrees > 10 in Supplemental Table 4. Furthermore, pathway enrichment was performed based on 220 macrophage-related DEGs in the PPI network (Figure 3(b)); nineteen KEGG pathways were emphasized, including cancer, cytokine-cytokine receptor interaction, biosynthesis of antibiotics, and purine metabolism pathways, among others.

**3.4. Screening of Macrophage-Related DEG Signatures to Build a Risk Prediction Model.** A total of 77 macrophage-related DEGs from the PPI network were found to be enriched in the KEGG pathways, and LASSO and RFE were implemented to further screen out DEG signatures to establish a diagnostic model with more accurate predictions. The parameter diagrams of LASSO and RFE are shown in Figure 4. Through LASSO regression analysis, 14 DEGs were found to have significant predictive advantages. Moreover, the RFE algorithm provided a combination of 20 DEGs as a stable prediction feature. Considering the intersection of the LASSO regression analysis and RFE algorithm, 10 macrophage-related DEG signatures (*EARS2*, *ATP2A2*, *COLGALT1*, *GART*, *WNT5A*, *AK5*, *ZBTB16*, *CCL17*, *ADORA3*, and *CXCR4*) were finally selected as an optimized gene set to predict AS diagnosis. Within the optimized gene set, *CXCR4*, *ZBTB16*, and *ADORA3* had relatively higher degrees of connection in the PPI network.



(a)



(b)

FIGURE 1: Continued.

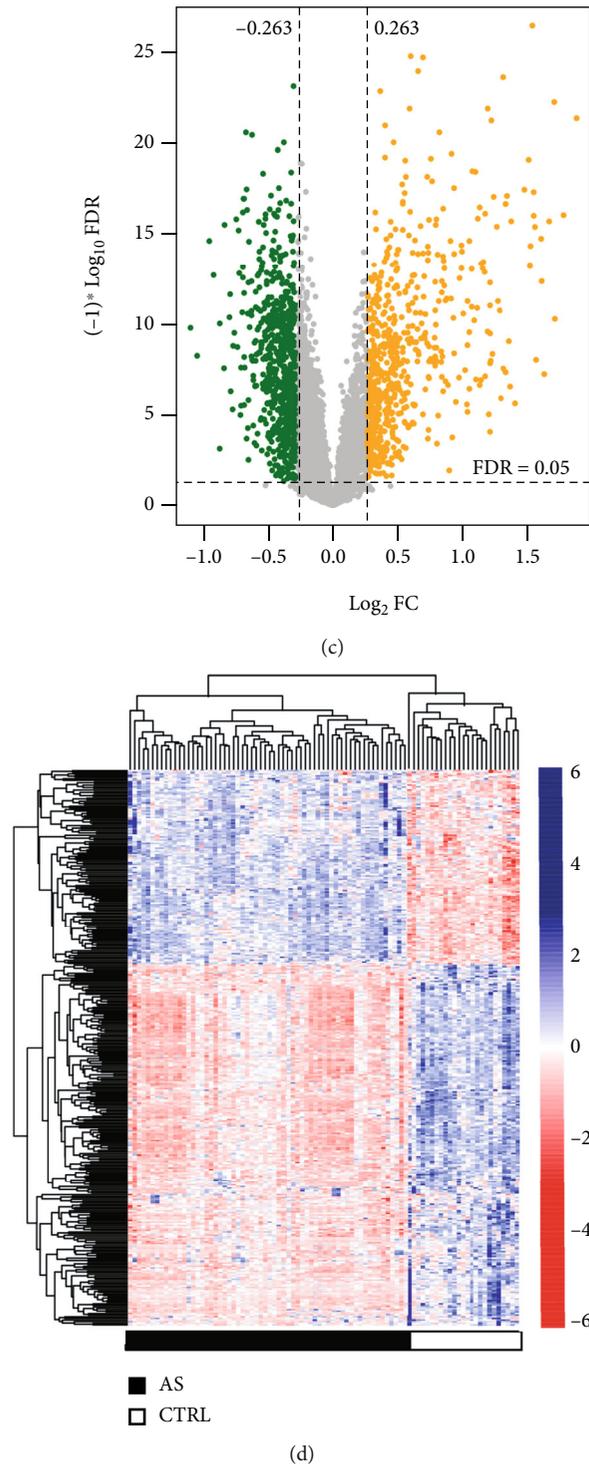
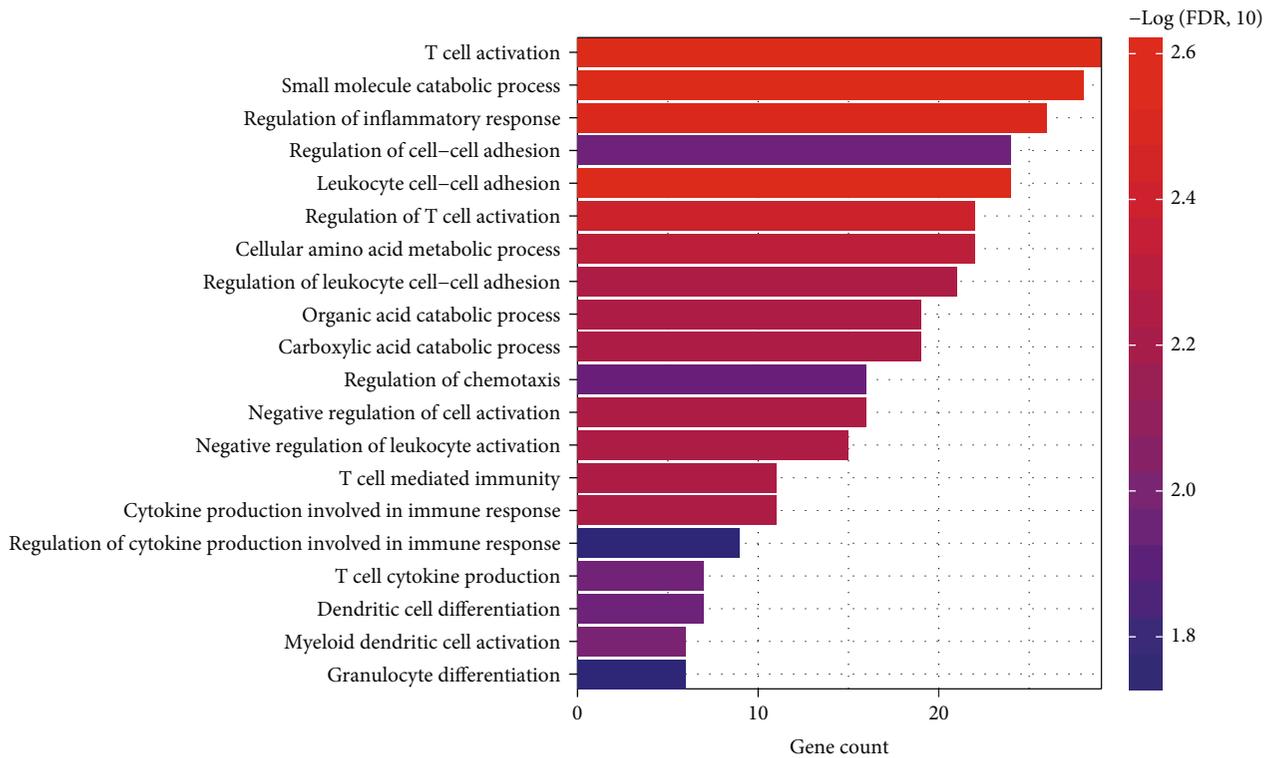


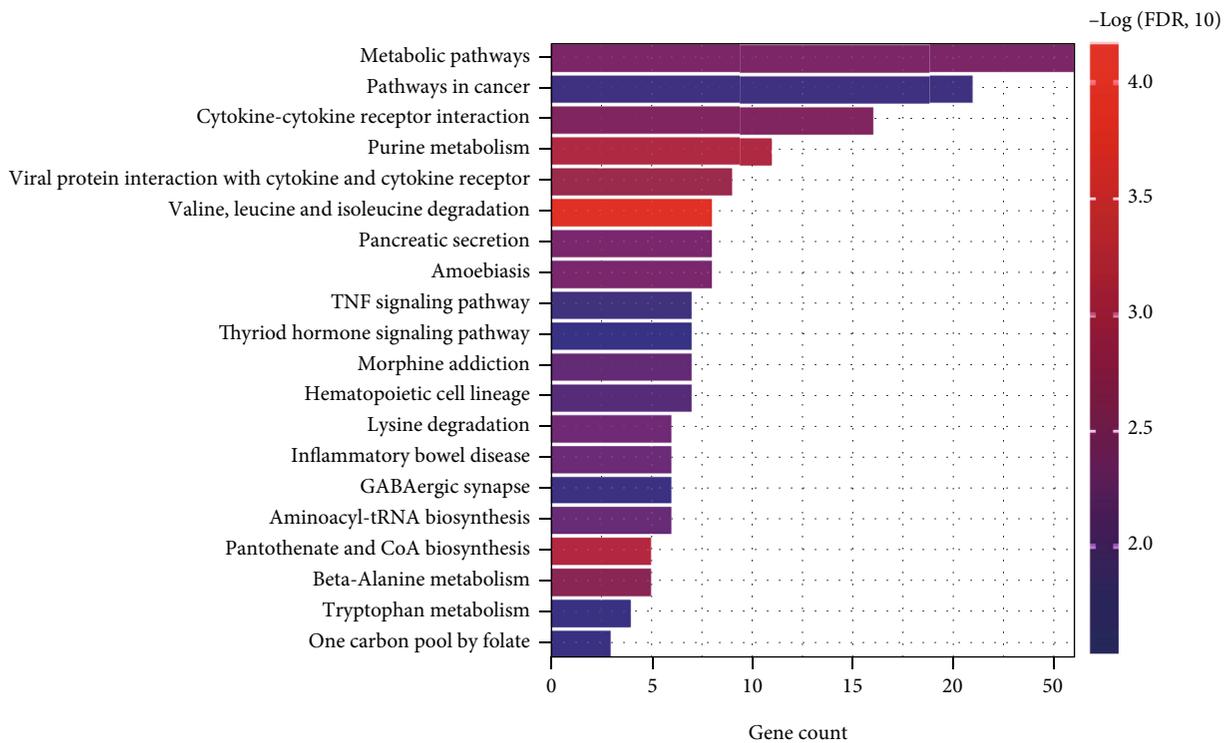
FIGURE 1: Screening of DEGs between AS and normal samples. (a, b) PCA diagrams before and after removing the batch effect on samples from the two datasets. Black and red dots represent samples in GSE137268 and GSE148004, respectively. (c) The volcano plot showed 1,189 DEGs between AS and the control groups screened from combined samples with the  $FDR < 0.05$  and  $|\log_2 FC| > 0.263$  criteria. The  $x$ -axis indicates the value of  $\log_2 FC$ , while the  $y$ -axis indicates the FDR value. The green and orange dots represent downregulated and upregulated DEGs, respectively, with statistical significance. (d) Heatmap based on the expression levels of DEGs.

3.5. Validation of Gene Expression and Prediction Model Efficacy. The expression data of the 10 macrophage-related DEG signatures were extracted and compared between AS samples and healthy controls in both the combined and

GSE112260 datasets. As shown in Figure 5(a), the expression of all 10 DEG signatures was significantly different between the two groups in the combined dataset. Meanwhile, in the GSE112260 dataset (Figure 5(b)), the expression trends of

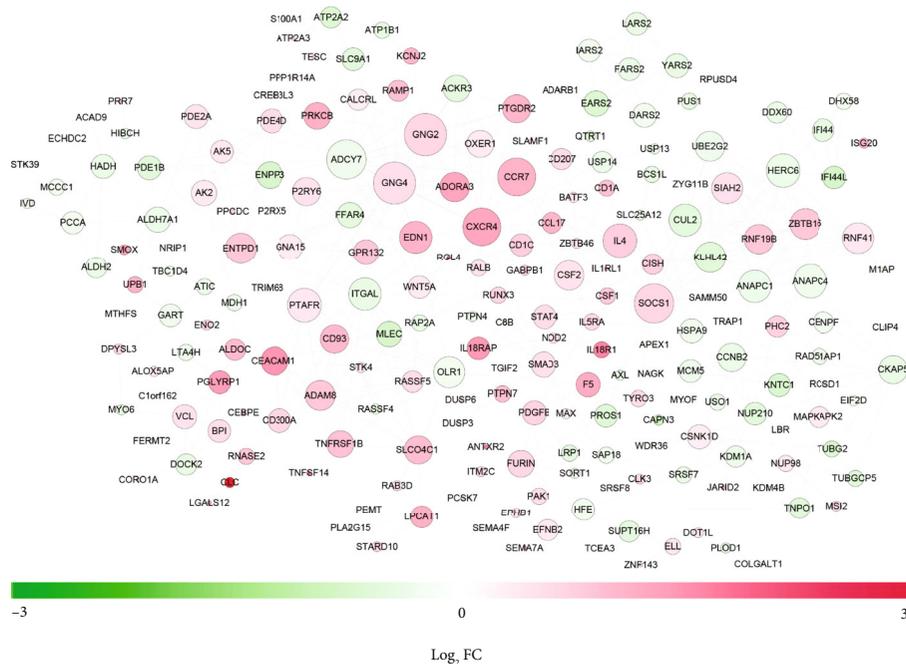


(a)

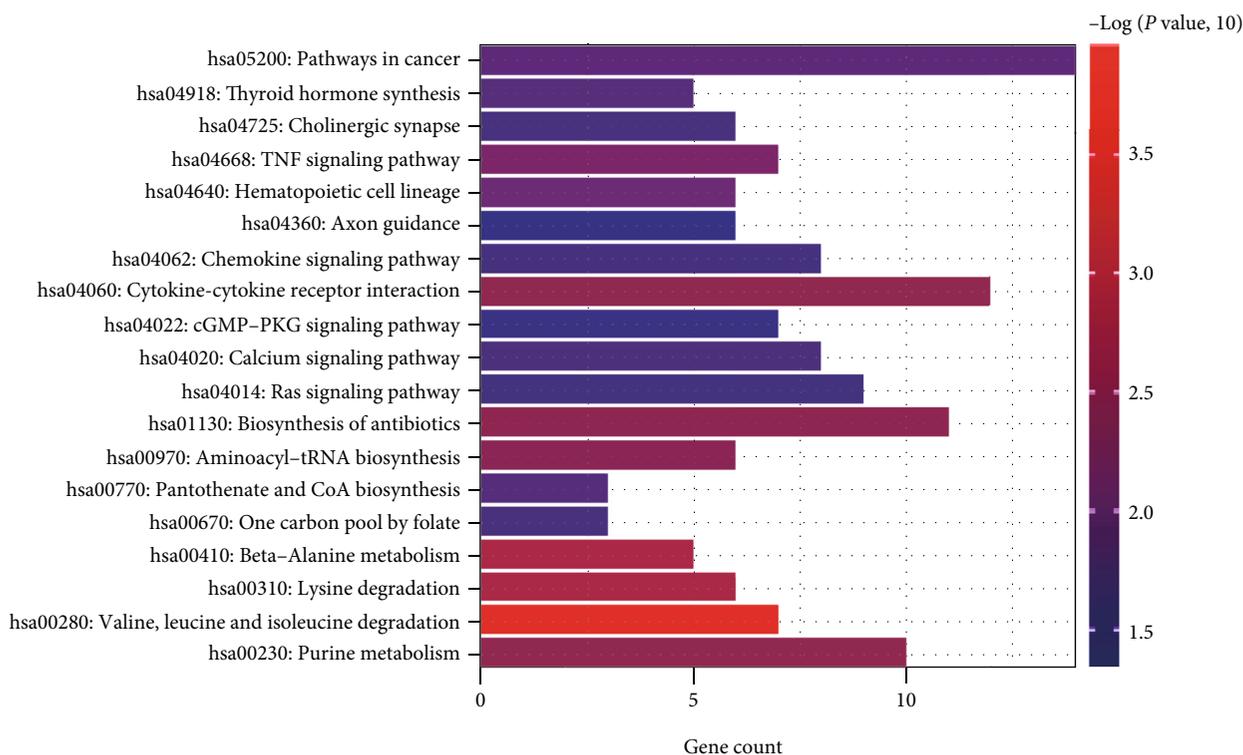


(b)

FIGURE 2: Functional analysis and pathway enrichment of macrophage-related DEGs. (a, b) Enriched GO-BP and KEGG pathways of macrophage-related DEGs with statistical significance. The  $x$ -axis indicates the gene count; the  $y$ -axis indicates the terms of biological processes or pathways, and the colors indicate the  $-\log P$  value.



(a)



(b)

FIGURE 3: PPI network construction and pathway enrichment of genes in the network. (a) Interactions between the coding proteins of the macrophage-related DEGs. The changes from green to red of nodes indicate the downregulation-upregulation of macrophage-related DEGs, and the sizes of nodes indicate the degree of connection in the network. (b) KEGG pathway enrichment of genes in the PPI network. The x-axis indicates the gene count; the y-axis indicates the terms of KEGG pathways, and the colors indicate the  $-\log P$  value.

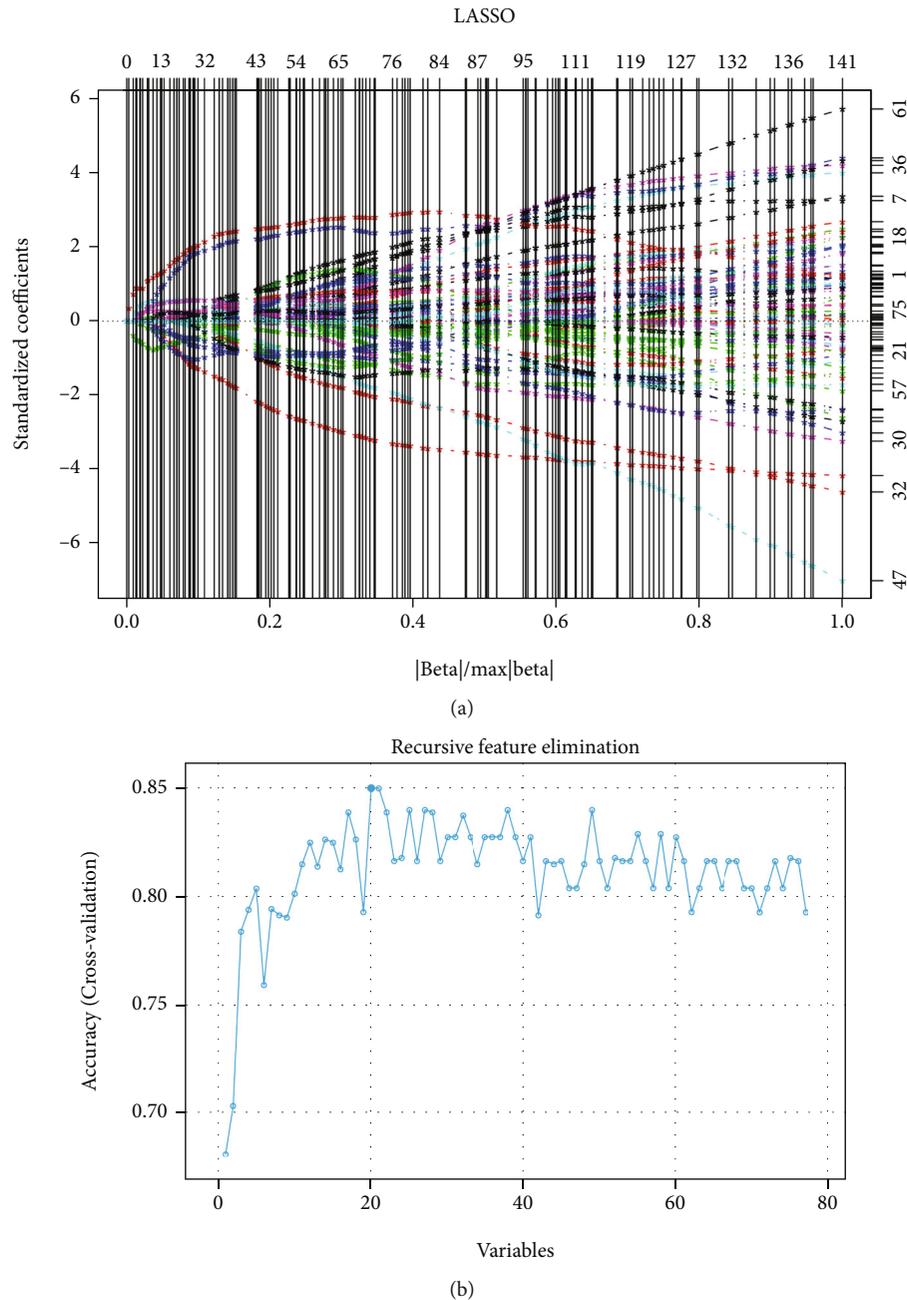


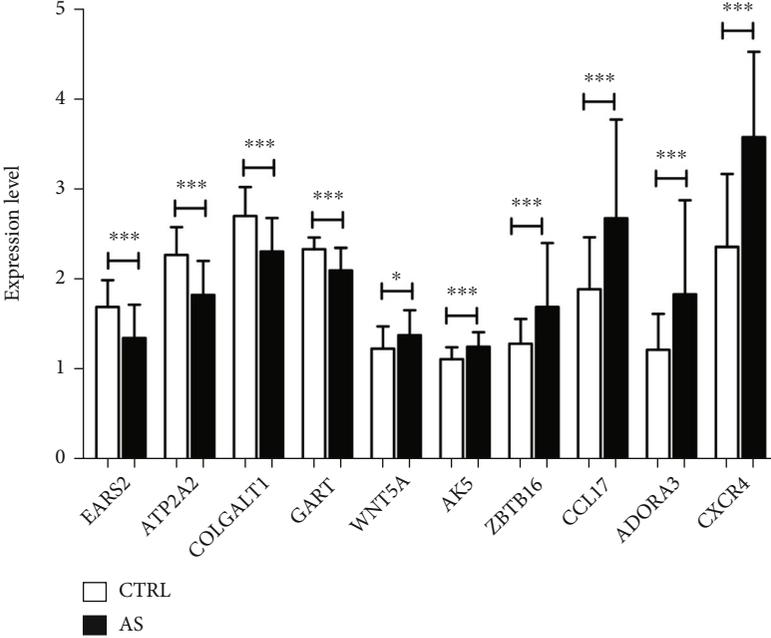
FIGURE 4: (a, b) Parameter diagrams of LASSO and RFE algorithms. (a) LASSO regression analysis provided variables with nonzero coefficients under different parameters. Below the  $x$ -axis are the values of different parameters in the combined dataset, and above the  $x$ -axis are the numbers of variables with the corresponding parameter. The left  $y$ -axis represents variable coefficients under different parameters, and the right  $y$ -axis represents the number of nonzero coefficient variables with the corresponding parameter. (b) The RFE algorithm provided a combination of variables with the highest accuracy of the prediction model. The  $x$ -axis indicates the number of variables, and the  $y$ -axis indicates the accuracy of the prediction model.

the 10 macrophage-related DEGs were in accordance with those in the combined dataset, and the expression differences of *EARS2*, *ATP2A2*, *WNT5A*, *CCL17*, *ADORA3*, and *CXCR4* were significant between the AS and control groups ( $P < 0.05$ ).

Based on the 10 macrophage-related DEG signatures, we constructed a diagnostic model using the combined dataset and validated model performance using the GSE112260 dataset. ROC curves showed excellent predic-

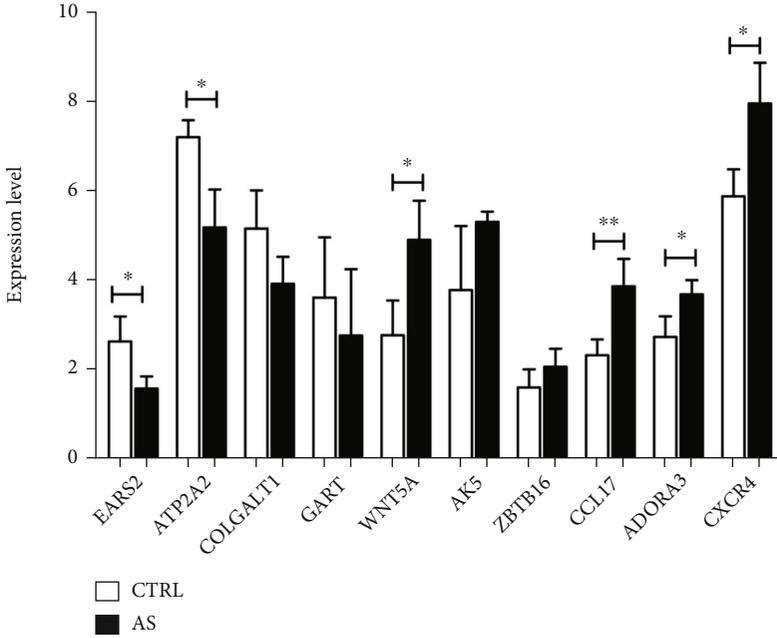
tive abilities of diagnostic models with an area under the curve (AUC) of 0.968 and 0.875 in the combined and GSE112260 datasets, respectively.

**3.6. Verification of 10 Differential Genes by qPCR in Clinical Samples.** The 10 macrophage-related DEG signatures (*EARS2*, *ATP2A2*, *COLGALT1*, *GART*, *WNT5A*, *AK5*, *ZBTB16*, *CCL17*, *ADORA3*, and *CXCR4*) were verified by qPCR. The results showed that the mRNA expressions of



(a)

GSE112260



(b)

FIGURE 5: Continued.

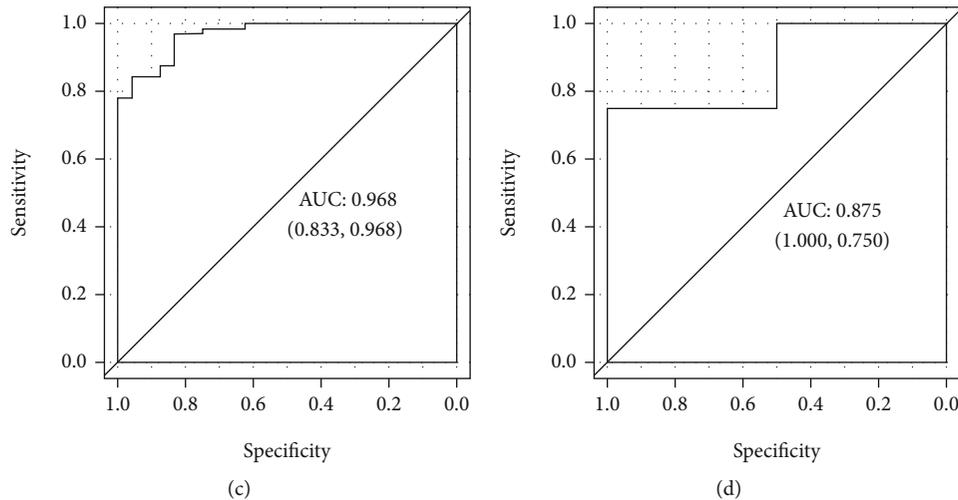


FIGURE 5: Expression validation of macrophage-related DEG signatures and performance validation of the diagnostic model. (a, b) Expression differences of 10 macrophage-related DEG signatures between the AS and control groups in the combined dataset and GSE112260. The  $x$ -axis indicates the 10 DEG signatures, and the  $y$ -axis indicates the expression level.  $*P < 0.05$ ,  $0.005 < **P < 0.05$ ,  $***P < 0.005$ . (c, d) ROC curves showing the abilities of diagnostic models in the combined dataset and GSE112260 with AUCs of 0.968 and 0.875, respectively. The  $x$ -axis indicates the value of specificity, and the  $y$ -axis indicates the value of sensitivity.

*EARS2*, *ATP2A2*, *COLGALT1*, and *GART* in the control group (healthy volunteers) were higher than in AS group, while the expressions of *WNT5A*, *AK5*, *ZBTB16*, *CCL17*, *ADORA3*, and *CXCR4* in control group were lower than that in the AS group (Figure 6).

#### 4. Discussion

AS is a multifaceted disease that affects all age groups, and genetic factors play an important role in the risk of developing AS. Although numerous genes, including *ORMDL3* and *GSDB* at locus 17q21 have been shown to contribute to the genetic etiology of AS [25, 26], there is a lack of strategies for integrating risk factors with multiple datasets to establish relationships among genetics, immunity, and AS. In this study, we first screened 1,189 DEGs from 63 AS and 24 control samples from the combined dataset. Macrophages were found to be significantly different between the two groups, and 439 macrophage-related DEGs were further screened after establishing the relationship between DEGs and the abundance of macrophage immune infiltration. These macrophage-related DEGs were mainly enriched in the GO-BP of immune and inflammatory responses as well as KEGG pathways of cytokine-cytokine receptor interaction and biosynthesis of antibiotics. By applying the LASSO and RFE algorithms, we finally screened 10 macrophage-related DEG signatures to establish a risk prediction model, which was shown to have excellent AS diagnostic abilities with AUCs of 0.968 and 0.875 in the ROC curves of the combined and GSE112260 datasets, respectively. Our findings propose significant biomarkers for AS diagnosis driven by macrophages.

AS is known to be associated with immune system activation, and both innate and adaptive immunity play roles in the immune mechanism of AS [27]. In this study, we found seven types of immune cells (including macrophages, natu-

ral killer cells, and immature dendritic cells, among others) that were significantly different between AS samples and healthy controls using a ssGSEA algorithm. Studies have shown that polarized M1/M2 macrophages function as antigen-presenting cells that may effectively activate Th1, Th2, Th17, or Treg cells in AS [28]. Moreover, a related study pointed out that the role of natural killer cells in the regression of peribronchial cell infiltration in AS may be to inhibit antigen-specific Th17 and Th1 immunity [29]. Additionally, dendritic cells are the most potent antigen-presenting cells in the immune system and play a central role in the allergen-driven Th2 immune response in AS [30]. This may explain the possible regulatory mechanism of macrophages with the combined action of natural killer cells and dendritic cells in activating Th1, Th2, and Th17 immune responses in AS attacks.

By evaluating the relationship between DEGs in AS and the proportion of macrophages, we obtained 439 macrophage-related DEGs that were mainly enriched in the biological processes of immune and inflammatory responses. Saradna et al. concluded that the regulation process of macrophages contained intricate interactions among various cytokines, chemokines, transcription factors, and immunomodulatory cells [11]; our results provide potential targets for such a complex interplay in macrophages and related inflammatory responses in AS. These macrophage-related DEGs were also enriched in the KEGG pathways of cytokine-cytokine receptor interactions. Zhao et al. found that the genes expressed in macrophages under hypoxic conditions were enriched in the cytokine-cytokine receptor interaction pathway [31]. Moreover, upregulated DEGs in acute respiratory distress syndrome were also determined to be associated with cytokine-cytokine receptor interactions [32]. Combined with the above findings, we hypothesized that dyspnea and hypoxia caused by asthma attacks might induce differential expression of these macrophage-

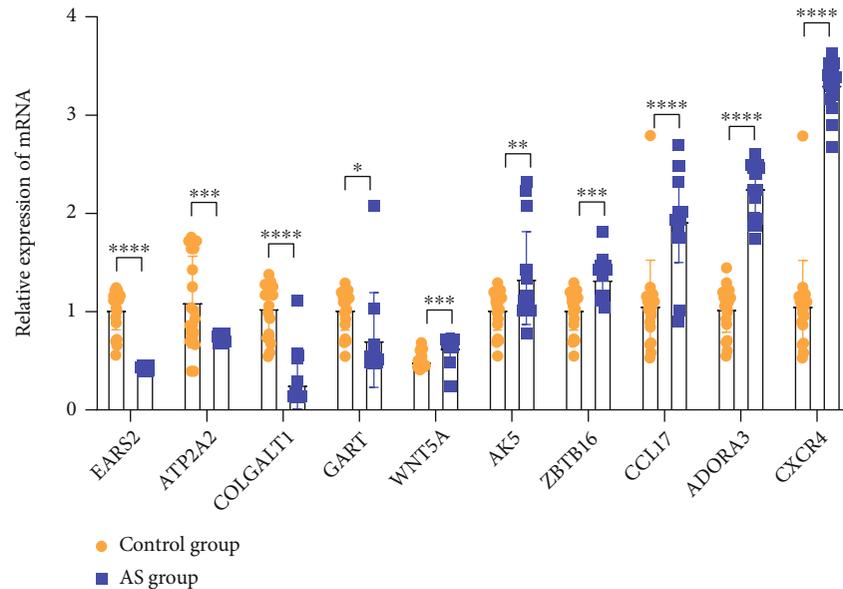


FIGURE 6: Verification of 10 differential genes by qPCR in clinical samples. The 10 macrophage-related DEG signatures (*EARS2*, *ATP2A2*, *COLGALT1*, *GART*, *WNT5A*, *AK5*, *ZBTB16*, *CCL17*, *ADORA3*, and *CXCR4*) were verified by qPCR. \* $P < 0.05$ ,  $0.005 < **P < 0.05$ ,  $***P < 0.005$ ,  $****P < 0.001$ .

related genes, which are closely related to the inflammatory and immune responses and cytokine-cytokine receptor interaction.

In the present study, LASSO and RFE algorithms were applied to establish a diagnostic model with more accurate predictions. LASSO analysis is a high-dimensional predictive regression method that integrates multiple biomarkers into a single model to enhance their predictive value [33]. In their prediction of miRNA-mRNA relationships in prostate cancer, Lu et al. identified LASSO as an informative tool in constructing diagnostic models with considerable advantages in sensitivity and specificity [34]. Kim SM and Kim Y also proposed that the prediction performance of LASSO regression for disease diagnosis was higher than that of stepwise logistic regression [35]. Additionally, SVM-RFE has been identified as an effective feature selection algorithm for feature screening in complex high-dimensional biological data and has been widely used in disease research and drug development [36]. Furthermore, Sanz et al. stated that in biomedical data analysis, RFE could accurately select variables and assess the direction and strength of associations [37]. The high specificity and sensitivity of the ROC curves in this study indicated that LASSO and RFE algorithms are superior in screening gene signatures to predict AS risk.

By taking the intersection of LASSO and RFE, we finally obtained 10 macrophage-related DEG signatures to predict AS risk and identify disease diagnostic performance. Among these candidate genes, the expression of *EARS2*, *ATP2A2*, *WNT5A*, *CCL17*, *ADORA3*, and *CXCR4* was further validated in both the combined and GSE112260 datasets. During the construction of a mouse model with conditional loss of function, Li et al. demonstrated that differentiation and migration of myofibroblasts were the main effects of *WNT5A* inactivation on alveolar formation [38]. Smooth

muscle-derived *WNT5A* enhanced Th2 inflammation in AS, leading to increased airway wall inflammation and remodeling [39]. Moreover, Williams et al. reported that mice with human rhinovirus-induced AS could upregulate the expression of *CCL17* by activating lung STAT6 [40]. However, Yuan et al. found that the deficiency of integrin  $\beta 4$  was involved in the Th2 response in allergic AS by down-regulating the *CCL17* pathway in airway epithelial cells [41]. The key roles of these feature genes in AS have been elucidated, but the molecular mechanism by which they mediate AS inflammation through macrophages is still unclear.

In this study, we screened gene signatures associated with macrophages and investigated their role in the diagnosis of AS. It is not only important for the diagnosis of asthma but also necessary to study the mechanism of asthma. It is important to further study the expression of these genes signatures in peripheral blood samples and tissue samples.

Although we found several interesting macrophage-related gene signatures and explored their roles in AS diagnosis, the inability to define the relationship between candidate genes and the severity and clinical AS phenotype caused by the lack of clinical information on samples was one of the limitations of the present study. In addition, the lack of multicenter large sample verification is also the deficiency of this study. In future studies, more solid tumor samples should be collected to verify the differences in the expression of these candidate genes and further explore their regulatory mechanisms in the inflammatory response and immune activation of AS.

## 5. Conclusion

Here, we obtained 439 DEGs in AS associated with macrophages, which also significantly differed in cell proportion

between AS and the controls. LASSO and RFE algorithms that can effectively identify and screen disease characteristic variables were then employed, and 10 macrophage-related DEG signatures were ultimately screened to establish a risk prediction model of AS by considering the intersection of relevant results. This prediction model showed excellent AS diagnostic abilities, with high sensitivity and specificity. Our findings may help to better understand the mechanisms of macrophage-mediated regulation in the pathogenesis of AS and provide potential diagnostic biomarkers for patients with AS.

## Abbreviations

AS:	Asthma
DEGs:	Differentially expressed genes
GEO:	Gene expression omnibus
PCA:	Principal component analysis
FDR:	Fold discovery rate
FC:	Fold change
GSEA:	Gene set enrichment analysis
ssGSEA:	Single sample gene set enrichment analysis
GSVA:	Gene set variation analysis
PCC:	Pearson correlation coefficient
GO-BP:	Gene ontology-biology process
KEGG:	Kyoto encyclopedia of genes and genomes
PPI:	Protein-protein interaction
LASSO:	Least absolute shrinkage and selection operator
RFE:	Recursive feature elimination
SVM:	Support vector machine
ROC:	Receiver operating characteristic
AUC:	Area under curve.

## Data Availability

The datasets generated during the current study are not publicly available but obtained from corresponding authors on reasonable request.

## Additional Points

**Highlights.** (1) We found 439 macrophage-related DEGs involved in the inflammation response of AS. (2) Macrophages showed significant differences in cell proportion between AS and controls. (3) LASSO and RFE were used to select gene signatures for building a risk prediction model. (4) Ten macrophage-related DEG signatures showed excellent AS diagnostic abilities. (5) The mRNA expressions of *EARS2*, *ATP2A2*, *COLGALT1*, and *GART* in the control group were higher than AS group, while the expressions of *WNT5A*, *AK5*, *ZBTB16*, *CCL17*, *ADORA3*, and *CXCR4* in the control group were lower than that in the AS group.

## Conflicts of Interest

The authors declare that no conflicts of interest exist.

## Authors' Contributions

All authors participated in the conception and design of the study. Qing Zhou and Furong Niu conceived the manuscript. Hong Shen and Xiaoshun Ai collected basic clinical information, processed the data, and performed the experiments. Yangyanqiu Wang and Jing Zhuang wrote the paper. Yani Zhou drew the figures. All authors read and approved the paper. Niu Furong and Zhou Qing contributed equally to this work and should be considered co-corresponding author.

## Acknowledgments

This work was supported by the Public Welfare Application Research Project of Huzhou Science and Technology Bureau (No. 2018GZ36) and Zhejiang Provincial Department of Science and Technology Project (No.LGF22H090037 and LGF22H290002). The authors gratefully acknowledge the database available to us for this study.

## Supplementary Materials

*Supplementary 1.* Supplemental Figure 1: the study flowchart.

*Supplementary 2.* Supplemental Figure 2: expression levels of samples (a) before and (b) after SVA batch effects.

*Supplementary 3.* Supplemental Figure 3: immune cells with significant differences in cell proportion between AS samples and controls. The *x*-axis indicates different kinds of immune cells with statistical significance, and the *y*-axis indicates the value of the cell fraction.

*Supplementary 4.* Supplemental Table 1: basic information of CRC samples.

*Supplementary 5.* Supplemental Table 2: primers of genes used for qPCR.

*Supplementary 6.* Supplemental Table 3: differences in proportion of twenty-eight types of immune cells between AS and controls.

*Supplementary 7.* Supplemental Table 4: nodes with degrees of connection over ten in the PPI network.

## References

- [1] M. I. Abdel-Aziz, A. H. Neerinx, S. J. Vijverberg, A. D. Kraneveld, and A. H. Maitland-van der Zee, "Omics for the future in asthma," in *Seminars in Immunopathology*, vol. 42, no. 1-pp. 111-126, Springer, Berlin Heidelberg, 2020.
- [2] A. Nanda and A. N. Wasan, "Asthma in adults," *The Medical Clinics of North America*, vol. 104, no. 1, pp. 95-108, 2020.
- [3] M. E. Kuruvilla, K. Vanijcharoenkarn, J. A. Shih, and F. E. Lee, "Epidemiology and risk factors for asthma," *Respiratory Medicine*, vol. 149, pp. 16-22, 2019.
- [4] S. A. G. Willis-Owen, W. O. C. Cookson, and M. F. Moffatt, "The genetics and genomics of asthma," *Annual Review of Genomics and Human Genetics*, vol. 19, no. 1, pp. 223-246, 2018.

- [5] F. Hoffmann, F. Ender, I. Schmutte et al., "Origin, localization, and Immunoregulatory properties of pulmonary phagocytes in allergic asthma," *Frontiers in Immunology*, vol. 7, p. 107, 2016.
- [6] V. Mishra, J. Banga, and P. Silveyra, "Oxidative stress and cellular pathways of asthma and inflammation: therapeutic strategies and pharmacological targets," *Pharmacology & Therapeutics*, vol. 181, pp. 169–182, 2018.
- [7] J. Balhara and A. S. Gounni, "The alveolar macrophages in asthma: a double-edged sword," *Mucosal Immunology*, vol. 5, no. 6, pp. 605–609, 2012.
- [8] M. Fricker and P. G. Gibson, "Macrophage dysfunction in the pathogenesis and treatment of asthma," *The European Respiratory Journal*, vol. 50, no. 3, p. 1700196, 2017.
- [9] T. A. van der Veen, L. E. S. de Groot, and B. N. Melgert, "The different faces of the macrophage in asthma," *Current Opinion in Pulmonary Medicine*, vol. 26, no. 1, pp. 62–68, 2020.
- [10] A. Mantovani, S. K. Biswas, M. R. Galdiero, A. Sica, and M. Locati, "Macrophage plasticity and polarization in tissue repair and remodelling," *The Journal of Pathology*, vol. 229, no. 2, pp. 176–185, 2013.
- [11] A. Saradna, D. C. Do, S. Kumar, Q. L. Fu, and P. Gao, "Macrophage polarization and allergic asthma," *Translational Research*, vol. 191, pp. 1–14, 2018.
- [12] T. Barrett, S. E. Wilhite, P. Ledoux et al., "NCBI GEO: archive for functional genomics data sets—update," *Nucleic Acids Research*, vol. 41, no. Database issue, pp. D991–D995, 2012.
- [13] E. E. Groth, M. Weber, T. Bahmer et al., "Exploration of the sputum methylome and omics deconvolution by quadratic programming in molecular profiling of asthma and COPD: the road to sputum omics 2.0," *Respiratory Research*, vol. 21, no. 1, 2020.
- [14] M. Paplińska-Goryca, K. Goryca, P. Misiukiewicz, P. Nejman-Gryz, K. Górka, and R. Krenke, "Genetic characterization of macrophages from induced sputum of patients with asthma and chronic obstructive pulmonary disease," *Polish archives of internal medicine*, vol. 128, no. 9, pp. 559–562, 2018.
- [15] J. T. Leek, W. E. Johnson, H. S. Parker, A. E. Jaffe, and J. D. Storey, "The sva package for removing batch effects and other unwanted variation in high-throughput experiments," *Bioinformatics (Oxford, England)*, vol. 28, no. 6, pp. 882–883, 2012.
- [16] M. E. Ritchie, B. Phipson, D. Wu et al., "Limma powers differential expression analyses for RNA-sequencing and microarray studies," *Nucleic Acids Research*, vol. 43, no. 7, article e47, 2015.
- [17] L. Ye, T. Zhang, Z. Kang et al., "Tumor-infiltrating immune cells act as a marker for prognosis in colorectal cancer," *Frontiers in Immunology*, vol. 10, p. 2368, 2019.
- [18] B. L. Li and X. P. Wan, "Prognostic significance of immune landscape in tumour microenvironment of endometrial cancer," *Journal of Cellular and Molecular Medicine*, vol. 24, no. 14, pp. 7767–7777, 2020.
- [19] D. Szklarczyk, J. H. Morris, H. Cook et al., "The STRING database in 2017: quality-controlled protein–protein association networks, made broadly accessible," *Nucleic acids research*, vol. 45, no. D1, pp. D362–d368, 2017.
- [20] P. Shannon, A. Markiel, O. Ozier et al., "Cytoscape: a software environment for integrated models of biomolecular interaction networks," *Genome Research*, vol. 13, no. 11, pp. 2498–2504, 2003.
- [21] M. G. Usai, A. Carta, and S. Casu, "Alternative strategies for selecting subsets of predicting SNPs by LASSO-LARS procedure," *BMC Proceedings*, vol. 6, no. S2, 2012.
- [22] T. M. Deist, F. Dankers, G. Valdes et al., "Machine learning algorithms for outcome prediction in (chemo)radiotherapy: an empirical comparison of classifiers," *Medical Physics*, vol. 45, no. 7, pp. 3449–3459, 2018.
- [23] Q. Wang and X. Liu, "Screening of feature genes in distinguishing different types of breast cancer using support vector machine," *Oncotargets and Therapy*, vol. 8, pp. 2311–2317, 2015.
- [24] X. Robin, N. Turck, A. Hainard et al., "pROC: an open-source package for R and S+ to analyze and compare ROC curves," *BMC Bioinformatics*, vol. 12, no. 1, p. 77, 2011.
- [25] S. Das, M. Miller, and D. H. Broide, "Chromosome 17q21 genes ORMDL3 and GSDMB in asthma and immune diseases," *Advances in Immunology*, vol. 135, pp. 1–52, 2017.
- [26] B. James, S. Milstien, and S. Spiegel, "ORMDL3 and allergic asthma: from physiology to pathology," *The Journal of Allergy and Clinical Immunology*, vol. 144, no. 3, pp. 634–640, 2019.
- [27] T. Boonpiyathad, Z. C. Sözen, P. Satitsuksanoa, and C. A. Akdis, "Immunologic mechanisms in asthma," *Seminars in Immunology*, vol. 46, article 101333, 2019.
- [28] Z. Jiang and L. Zhu, "Update on the role of alternatively activated macrophages in asthma," *Journal of asthma and allergy*, vol. 9, pp. 101–107, 2016.
- [29] M. M. Gorska, "Natural killer cells in asthma," *Current Opinion in Allergy and Clinical Immunology*, vol. 17, no. 1, pp. 50–54, 2017.
- [30] I. Morianos and M. Semitekolou, "Dendritic cells: critical regulators of allergic asthma," *International Journal of Molecular Sciences*, vol. 21, no. 21, p. 7930, 2020.
- [31] Y. Zhao, B. Zhao, X. Wang et al., "Macrophage transcriptome modification induced by hypoxia and lactate," *Experimental and Therapeutic Medicine*, vol. 18, no. 6, pp. 4811–4819, 2019.
- [32] L. Liao and P. Liao, "Bioinformatics analysis of the potential biomarkers for acute respiratory distress syndrome," *Bioscience Reports*, vol. 40, no. 9, 2020.
- [33] Y. Jiang, Q. Zhang, Y. Hu et al., "ImmunoScore Signature," *Annals of Surgery*, vol. 267, no. 3, pp. 504–513, 2018.
- [34] Y. Lu, Y. Zhou, W. Qu, M. Deng, and C. Zhang, "A lasso regression model for the construction of microRNA-target regulatory networks," *Bioinformatics (Oxford, England)*, vol. 27, no. 17, pp. 2406–2413, 2011.
- [35] S. M. Kim, Y. Kim, K. Jeong, H. Jeong, and J. Kim, "Logistic LASSO regression for the diagnosis of breast cancer using clinical demographic data and the BI-RADS lexicon for ultrasonography," *Ultrasonography*, vol. 37, no. 1, pp. 36–42, 2018.
- [36] X. Lin, C. Li, Y. Zhang, B. Su, M. Fan, and H. Wei, "Selecting feature subsets based on SVM-RFE and the overlapping ratio with applications in bioinformatics," *Molecules Basel, Switzerland*, vol. 23, no. 1, 2018.
- [37] H. Sanz, C. Valim, E. Vegas, J. M. Oller, and F. Reverter, "SVM-RFE: selection and visualization of the most relevant features through non-linear kernels," *BMC Bioinformatics*, vol. 19, no. 1, p. 432, 2018.
- [38] C. Li, S. M. Smith, N. Peinado et al., "WNT5a-ROR signaling is essential for alveologenesis," *Cell*, vol. 9, no. 2, p. 384, 2020.

- [39] T. Koopmans, L. Hesse, M. C. Nawijn et al., “Smooth-muscle-derived WNT5A augments allergen-induced airway remodeling and Th2 type inflammation,” *Scientific Reports*, vol. 10, no. 1, p. 6754, 2020.
- [40] T. C. Williams, D. J. Jackson, S. Maltby et al., “Rhinovirus-induced CCL17 and CCL22 in asthma exacerbations and differential regulation by STAT6,” *American Journal of Respiratory Cell and Molecular Biology*, vol. 64, no. 3, pp. 344–356, 2021.
- [41] L. Yuan, X. Zhang, M. Yang et al., “Airway epithelial integrin  $\beta$ 4 suppresses allergic inflammation by decreasing CCL17 production,” *Clinical Science (London, England)*, vol. 134, no. 13, pp. 1735–1749, 2020.