DATABASE ANALYSIS

e-ISSN 1643-3750
© Med Sci Monit, 2026; 32: e952390
DOI: 10.12659/MSM.952390

Received: 2025.12.10
Accepted: 2026.04.26
Available online: 2026.06.04
Published: 2026.XX.XX

Identification of *EGR2* and *TGFB1* as Potential Immunoregulatory Biomarkers for Allergic Rhinitis: An Exploratory Bioinformatics and Pilot Validation Study

Authors' Contribution:
Study Design A
Data Collection B
Statistical Analysis C
Data Interpretation D
Manuscript Preparation E
Literature Search F
Funds Collection G

ABE 1-3 **Yisen Liu**
BE 3 **Shuo Guo**
CE 4 **Feng Cao**
AEF 1,2 **Xixi Chen**
ABE 1,2 **Yehai Liu**

1 Department of Otolaryngology, Head and Neck Surgery, First Affiliated Hospital of Anhui Medical University, Hefei, Anhui, PR China
2 Department of Allergy, First Affiliated Hospital of Anhui Medical University, Hefei, Anhui, PR China
3 Department of Otolaryngology, Head and Neck Surgery, Third People's Hospital of Hefei, Hefei, Anhui, PR China
4 Department of Otolaryngology, Head and Neck Surgery, Second People's Hospital of Hefei, Hefei, Anhui, PR China

Corresponding Author: Yehai Liu, No. 203, Jixi Road, Shushan District, Hefei, Anhui, China, Phone: +8613856906986, e-mail: liuyehai@ahmu.edu.cn
Financial support: This study was supported by the National Natural Science Foundation (Grant No. 82171127)
Conflict of interest: None declared

Background: Allergic rhinitis (AR) is a common inflammatory disorder of the nasal mucosa triggered by allergens and characterized by sneezing, nasal congestion, and itching. This study aimed to identify and validate potential biomarkers associated with immune dysregulation in AR via bioinformatics analyses initiated with chemokine-related genes.

Material/Methods: AR-related datasets (GSE75011 and GSE50223) and 69 chemokine-related genes were analyzed via differential expression analysis and weighted gene co-expression network analysis to identify candidate genes. Random forest analysis was used to select optimal features. Expression analysis identified potential immunoregulatory biomarkers that were downregulated in AR samples. A nomogram incorporating these biomarkers was constructed as an exploratory tool to estimate individual AR probability and support future diagnostic model development.

Results: Early growth response protein 2 (*EGR2*) and transforming growth factor-β1 (*TGFB1*) were identified as chemokine-related biomarkers in AR. Immune infiltration analysis showed significant correlations between *TGFB1* and M2 macrophages, as well as activated memory CD4+ T cells. Drug prediction identified tretinoin as a potential therapeutic agent targeting both biomarkers; molecular docking confirmed stable binding of the 3 top-ranked predicted drugs with *TGFB1*. Pilot reverse transcription-quantitative polymerase chain reaction analysis demonstrated downregulation of *EGR2* and *TGFB1* in AR samples, providing preliminary supportive evidence. Gene set enrichment analysis indicated involvement in the MAPK and p53 signaling pathways, suggesting roles in cell proliferation, apoptosis, cell cycle regulation, and stress responses.

Conclusions: *EGR2* and *TGFB1* offer potential immunoregulatory biomarkers for AR diagnosis and may provide preliminary insights into AR pathogenesis and potential treatment strategies.

Keywords: biomarkers • chemokines • rhinitis, allergic

Full-text PDF: https://www.medscimonit.com/abstract/index/idArt/952390

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]

Introduction

Allergic rhinitis (AR) is a noninfectious inflammatory disorder of the nasal mucosa mediated by specific immunoglobulin E (IgE). It is triggered by abnormal IgE production in response to allergens, leading to activation of mast cells and eosinophils, with subsequent release of mediators (eg, histamine) that contribute to clinical manifestations including sneezing, nasal congestion, and itching [1]. Epidemiological studies have shown that AR affects 20% to 30% of adults and up to 40% of children, causing varying degrees of negative impact on quality of life [2]. Patients with AR often experience reduced academic and work performance, impaired sleep, and diminished quality of life; they may even develop psychological disorders such as depression, resulting in a substantial socioeconomic burden [3].

Historically, AR diagnosis has primarily relied on clinical history. Recently, multiple strategies have been utilized to explore the pathological mechanisms and therapeutic targets of AR, including conventional pharmacological studies, emerging network pharmacology approaches, and bioinformatics-based screening with high-throughput data [4,5]. Despite these advances, challenges persist in achieving high diagnostic specificity and accuracy [6]. Current treatments for AR, including antihistamines, corticosteroids, and leukotriene receptor antagonists, can rapidly relieve symptoms but often have limited long-term efficacy; they are associated with adverse effects such as epistaxis and drowsiness. Additionally, these treatments may be ineffective in 30% to 60% of patients with AR. Although allergen immunotherapy can target the underlying mechanisms involved in AR development and progression, its clinical application is often hindered by practical challenges such as prolonged treatment duration and high cost [5,7,8]. Thus, the identification of novel biomarkers is crucial to improve diagnostic accuracy and enable personalized treatment strategies for AR.

Chemokines are a class of small, secreted proteins with molecular masses of approximately 7 to 14 kDa. Based on the number and position of cysteine residues in their amino acid sequences, they are classified into 4 subfamilies: C, CC, CXC, and CX3C [9]. Chemokines regulate the directed migration and activation of leukocytes during inflammation and homeostasis in a time- and site-dependent manner [10]. Chemokines play a critical role in recruiting immune cells to inflamed tissues, thereby facilitating pathogen elimination and tissue repair during infection or injury [11]. Under inflammatory conditions, chemokine imbalance has been associated with the pathogenesis of various inflammatory diseases, including cardiovascular diseases, arthritis, and neuropathic pain. The uncontrolled inflammatory state resulting from chemokine dysregulation remains a major challenge in clinical management [12-14]. Chemokines and their receptors have also been implicated in the sensitization phase, as well as the early and late responses of AR, by promoting inflammatory cell recruitment, differentiation, and release of allergic mediators [15]. Therefore, chemokines may play an important role in AR pathogenesis, although the specific pathways and molecular regulatory mechanisms involving chemokine-related genes (CRGs) remain unclear.

In summary, chemokines and their related genes play a crucial role in the inflammatory recruitment processes involved in AR; however, the upstream regulatory network and key immune regulatory hubs have not been elucidated. Currently, clinical management of AR faces 2 major challenges: a lack of highly specific biomarkers of immune homeostasis for diagnosis and the limited ability of existing therapeutic approaches to correct upstream immune dysregulation [16,17]. These challenges highlight the substantial translational value of identifying and validating key genes associated with immune regulatory dysfunction as potential biomarkers and therapeutic targets.

To address this knowledge gap, we conducted an integrative bioinformatics investigation. Beginning with CRGs, we analyzed publicly available AR gene expression datasets to identify chemokine-associated biomarkers that may function as immune regulatory hubs. We subsequently applied a comprehensive approach integrating machine-learning-based screening, diagnostic model construction, immune infiltration analysis, and drug prediction to perform multidimensional validation and functional characterization of these biomarkers. This study aims not only to identify novel biomarkers but also to provide new insights and potential targets for future AR precision diagnosis and treatment strategies based on immune homeostasis restoration.

Material and Methods

Data Gathering and Preprocessing

Gene expression datasets related to AR—GSE75011 (platform: GPL16791) and GSE50223 (platform: GPL6884)—were obtained from the Gene Expression Omnibus (GEO; http://www.ncbi.nlm.nih.gov/geo/) database. Both datasets were derived from peripheral blood mononuclear cells, making them suitable for studies of systemic immune regulation. GSE75011 contains RNA-seq data from 25 patients with AR and 15 healthy controls, whereas GSE50223 contains microarray data from 21 patients with AR and 21 healthy controls.

For the RNA-seq dataset GSE75011, raw count data were processed using the DESeq2 package (ver. 1.42.0). To reduce statistical noise introduced by low-expression genes, genes with a total raw count of 1 or less across all samples were removed. After filtering, the remaining genes were converted to human gene symbols. When multiple Ensembl identifiers corresponded

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]

to the same gene symbol, only the entry with the highest total raw count was retained.

For the microarray dataset GSE50223, a pre-normalized expression matrix was downloaded. Probe annotations were updated by matching GPL6884 probe identifiers with current gene symbols. For genes represented by multiple probes, the probe with the highest average expression across all samples was selected as the representative. Given the pivotal role of chemokines and their receptors in orchestrating immune cell recruitment during allergic inflammation, we retrieved 69 CRGs (such as *CXCL1*, *CXCL2*, *CXCL3*, and *CCL1*) from published literature [18] as a strategic starting point for bioinformatics screening to identify key regulators potentially involved in AR pathogenesis.

Acquisition of Differentially Expressed Genes (DEGs)

DEGs between AR and control samples in the GSE75011 dataset were identified using the R package DESeq2 (ver. 1.42.0) [19]. To prioritize potentially relevant genes for exploratory analysis, thresholds of $|\log_2$ fold change (FC)$| > 0.5$ and raw $P < 0.05$ were applied. These criteria were selected to balance sensitivity and specificity in a hypothesis-generating context. The R package ggplot2 (ver. 3.4.4) [19] was used to generate a volcano plot for visualization of DEGs. Additionally, a heat map illustrating DEG expression patterns was generated using the R package ComplexHeatmap (ver. 2.14.0) [20].

Weighted Gene Co-Expression Network Analysis (WGCNA)

Based on the expression profiles of CRGs, the single-sample gene set enrichment analysis (ssGSEA) algorithm implemented in the GSVA package (ver. 1.46.0) [21] was used to calculate ssGSEA scores for CRGs. Differences in these scores between AR and control samples were assessed via the Wilcoxon rank-sum test (threshold: $P < 0.05$). Subsequently, using ssGSEA scores as traits, the WGCNA R package (ver. 1.42.0) [22] was utilized to construct a weighted gene co-expression network for identifying key module genes most strongly associated with CRGs.

Euclidean distance-based hierarchical clustering was performed for all samples in the GSE75011 dataset to identify and remove potential outliers. Determination of the optimal soft-thresholding power (β) was essential to ensure that the co-expression network conformed to a scale-free topology. The β value was selected under the criteria that the scale-free topology fit index (R^2) exceeded 0.8 and mean connectivity approached 0. Based on the selected β value, co-expression modules were identified using a hybrid dynamic tree-cutting algorithm, with the minimum module size set to 100 genes, a deepSplit parameter of 4, and a module merging threshold of 0.2. The deepSplit parameter controlled clustering sensitivity to ensure sufficient module size while preserving strong co-expression relationships among genes within each module. The module merging threshold enabled integration of highly correlated modules, reducing redundancy while retaining biologically relevant modules. Subsequently, correlation coefficients between each co-expression module and ssGSEA scores were calculated. The module exhibiting the strongest absolute correlation with the ssGSEA score was designated as the key module (thresholds: $|cor| > 0.3$, $P < 0.05$); genes within this module were defined as key module genes.

Identification and Functional Analyses of Candidate Genes

Candidate genes associated with CRGs and AR were identified by intersecting key module genes with DEGs. To elucidate the potential functions of these candidate genes, Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analyses [23,24] were performed using the clusterProfiler package in R (ver. 4.7.1.003) [25]. These analyses were conducted to identify associated biological processes, molecular functions, cellular components, and KEGG pathways (threshold: $P < 0.05$). Additionally, interactions among proteins encoded by the candidate genes were analyzed through a protein-protein interaction (PPI) network. Network visualization was performed using Cytoscape software (ver. 3.9.1) [26], and candidate genes were submitted to the STRING database (https://string-db.org/) using an interaction score threshold greater than 0.9. Candidate genes included in the PPI network were subsequently ranked according to Degree score; the top 20 genes were selected for subsequent machine learning analysis.

Discovery of Potential Biomarkers in AR

Three machine learning classifiers—generalized linear model (GLM), random forest (RF; parameter settings: ntree = 23, mtry = 2), and support vector machine (SVM; method = svmLinear)—were trained using the GSE75011 dataset. To reduce the risk of overfitting and evaluate model stability, bootstrap resampling with 25 replications, implemented in the caret R package (ver. 6.0-93) [27], was used for internal validation of the GLM and SVM models. For the RF model, performance was evaluated via built-in out-of-bag error estimation. Subsequently, the root mean square of residuals (RMSR) and root mean square error (RMSE) of each model were calculated based on the full GSE75011 dataset, and receiver operating characteristic (ROC) curves were generated to calculate the area under the curve (AUC). ROC curves were plotted using the R package pROC (ver. 2.4.3) [28]. The classifier with the lowest RMSR, lowest RMSE, and highest AUC in this exploratory internal validation setting was selected for further analysis. Given the absence of external validation, all classifier results were considered exploratory. Genes included in the selected classifier were

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]

regarded as candidate biomarkers. Furthermore, confusion matrices, accuracy, precision, recall, and F1 scores for the GLM, RF, and SVM models were generated using the confusionMatrix function in the caret package (ver. 6.0-93) to further evaluate model performance. The RF model was constructed using the randomForest R package (ver. 4.7-1.1) [29]. Potential biomarkers were subsequently identified through expression analysis. Candidate biomarkers displaying significantly different expression between AR and control samples in both the GSE75011 and GSE50223 datasets, with fully consistent expression trends across the 2 datasets, were defined as potential biomarkers (threshold: $P < 0.05$).

Nomogram Establishment

To explore the potential utility of the identified biomarkers for estimating AR probability within the discovery dataset, a nomogram was constructed using the rms package in R (ver. 6.5-0) [30]. In the nomogram, each factor (ie, each biomarker) was assigned a score, and the sum of these scores corresponded to the total score. A higher total score indicated a greater probability of developing AR. To evaluate predictive performance, calibration curves were generated using the calibrate function in the rms package (ver. 6.5-0). The Hosmer-Lemeshow test was also performed (threshold: $P > 0.05$). A calibration curve slope closer to 1 indicated greater predictive accuracy of the nomogram. Additionally, an ROC curve was generated using the pROC package (ver. 2.4.3), and the AUC was calculated to further assess predictive performance (threshold: AUC > 0.7).

Gene Set Enrichment Analysis (GSEA)

GSEA was subsequently performed using the clusterProfiler R package to identify pathways associated with the biomarkers. Samples in the GSE75011 dataset were divided into high- and low-expression groups according to median biomarker expression level, then subjected to differential expression analysis. Log_2FC values were calculated for each DEG, and genes were ranked in descending order according to log_2FC values. The ranked gene list was used for GSEA with c2.cp.kegg.v7.4.symbols.gmt as the reference gene set obtained from the Molecular Signatures Database (MSigDB, https://www.gsea-msigdb.org/) (threshold: $P < 0.05$).

Immunoinfiltration Characterization

To characterize the immune cell infiltration landscape during AR progression, the CIBERSORT algorithm implemented in the Immunedeconv R package (ver. 2.0.4) [31] was used to estimate the proportions of 22 immune cell types in each sample from the GSE75011 dataset. To ensure reliability, only samples with $P < 0.05$ based on confidence calculations were retained. Differentially infiltrated immune cells (DICs) between AR and control samples were subsequently identified (threshold: $P < 0.05$). Spearman correlation analysis was performed for exploratory hypothesis generation using thresholds of $|cor| > 0.3$ and raw $P < 0.05$ to identify potentially relevant associations among DICs, as well as between DICs and each biomarker.

Drug Prediction and Molecular Docking

To identify potential therapeutic agents targeting biomarkers for AR treatment, biomarkers were submitted to the Drug-Gene Interaction Database (DGIdb; https://www.dgidb.org) to identify candidate drugs, and a drug-biomarker interaction network was visualized using Cytoscape software. Drugs were ranked according to interaction scores; the top 3 candidates were selected for molecular docking to investigate their interaction patterns with target biomarkers. The chemical structures of the drugs and 3-dimensional structures of the biomarkers were obtained from the PubChem (https://pubchem.ncbi.nlm.nih.gov/) and Protein Data Bank (PDB; https://www.rcsb.org/) databases, respectively. Molecular docking was performed and visualized using the online platform CB-Dock 1 (https://cadd.labshare.cn/cb-dock/). Prior to formal docking, CB-Dock 1 automatically performed energy minimization according to protein and ligand input. Binding energies below -5 kcal/mol were considered indicative of strong binding affinity between the biomarker and drug.

Reverse Transcription-Quantitative Polymerase Chain Reaction (RT-qPCR)

To preliminarily assess the mRNA expression levels of potential biomarkers, whole blood samples were collected from 5 patients with AR and 5 healthy controls for RT-qPCR analysis. Whole blood samples were selected due to their clinical accessibility, noninvasive nature, and practicality for routine diagnostic use. Sample collection criteria and clinical characteristics are presented in **Supplementary Table 1**. All samples were collected from the Third People's Hospital of Hefei, and informed consent was obtained from all participants. This study was conducted in accordance with the Declaration of Helsinki and was approved by the Ethics Committee of the Third People's Hospital of Hefei (approval number: 2024LLWL038). To protect participant privacy, all samples were de-identified immediately after collection and labeled using anonymous codes only (eg, AR patient 1, Control 1). Personally identifiable information (eg, names, identification numbers, and hospitalization numbers) was excluded from the study database and manuscript. Investigators could not trace sample codes back to individual participants, thereby ensuring ethical compliance in data handling and privacy protection.

The RT-qPCR procedure began with extraction of total RNA using TRIzol reagent (Ambion, USA), in accordance with the manufacturer's protocol. RNA concentration and quality were rapidly

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]

Table 1. Primer sequences used for reverse transcription quantitative polymerase chain reaction.

Primer	Sequence
EGR2_F	AACACAGACAGGAGAGAGTCAG
EGR2_R	GCGGTCATCATTTGCTCCTC
TGFB1_F	TCAACGCAGGGTTCACTACC
TGFB1_R	GAAGTTGGCATGGTAGCCCT
GAPDH_F	CGAAGGTGGAGTCAACGGATTT
GAPDH_R	ATGGGTGGAATCATATTGGAAC

assessed using a NanoPhotometer N50. Subsequently, cDNA synthesis was performed using the SureScript First Strand cDNA Synthesis Kit (Servicebio, China). For qPCR analysis, each sample (biological replicates: 5 AR samples and 5 control samples) was analyzed in triplicate (technical replicates) to ensure reproducibility. Glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*) served as the internal reference gene, and relative mRNA expression levels were measured using a CFX Connect Thermal Cycler (Bio-Rad, USA). Differences between AR and control groups were calculated using the $2^{-\Delta\Delta CT}$ method [32]. Details of all primers are provided in **Table 1**. Given the small sample size, RT-qPCR data were compared between groups using the nonparametric Mann-Whitney U test (threshold: raw $P < 0.05$).

Statistical Analysis

All statistical analyses were performed using R software (ver. 4.2.2). For comparisons between 2 groups (eg, AR vs control), the Wilcoxon rank-sum test was applied, and raw *P* values were reported. Given the exploratory nature of this study, no adjustments for multiple testing were applied to any analyses, including differential expression analysis, GO/KEGG enrichment, GSEA, immune infiltration analysis, machine learning feature selection, and RT-qPCR validation. Therefore, all findings should be interpreted as hypothesis-generating and require validation in independent cohorts. Raw *P* values < 0.05 were considered statistically significant. The overall analytical workflow is illustrated in **Supplementary Figure 1**. Detailed sample statistics for each analytical step are provided in **Supplementary Table 2**.

Results

Screening and Enrichment Analysis of 85 Candidate Genes

Differential expression analysis identified 483 DEGs in AR samples from the GSE75011 dataset, including 157 upregulated and 326 downregulated genes ($|\log2FC| > 0.5$, $P < 0.05$). The top 10 upregulated and top 10 downregulated DEGs, ranked according to $\log_2$FC values, were highlighted in the volcano plot (**Figure 1A**). A heat map was constructed to illustrate the expression patterns of these top 10 upregulated and downregulated DEGs (**Figure 1B**). The ssGSEA score for CRGs was significantly lower in the AR group than in the control group ($P = 0.0023$) (**Figure 2**). Hierarchical clustering analysis indicated no outliers or abnormal samples in the GSE75011 dataset (**Figure 1C**). The optimal soft-thresholding power (β) was calculated as 16; at this value, the scale-free topology fit index (R^2) reached 0.8010 and mean connectivity approached 0 (**Figure 1D**). In total, 6 co-expression modules were identified, excluding the MEgray module, which contained unclassified genes (**Figure 1E**). Based on the predefined criteria of $|cor| > 0.3$ and $P < 0.05$, the MEred module exhibited the strongest positive correlation with CRG scores ($cor = 0.41$, $P = 0.008$) and was therefore selected as the key module. This module contained 190 genes (**Figure 1F**).

In total, 85 candidate genes were identified by intersecting the 190 key module genes with the 483 DEGs (**Figure 3A**). Functional analysis showed that GO enrichment identified 707 significant terms ($P < 0.05$), including 13 cellular component, 54 molecular function, and 640 biological process terms. These enriched terms included negative regulation of phosphorylation, metabolic processes, phosphorus metabolic processes, protein phosphorylation, synaptic vesicles, and GTPase activity, among others. Additionally, candidate genes were enriched in 53 KEGG pathways ($P < 0.05$). The top 15 significant pathways, ranked according to P value, included the mitogen-activated protein kinase (MAPK) signaling pathway, tumor necrosis factor signaling pathway, interleukin (IL)-17 signaling pathway, and apoptosis (**Figure 3B**). Overall, these findings suggest that candidate genes play important roles in cell signaling, gene regulation, and enzyme activity. A PPI network was constructed by submitting the 85 candidate genes to the STRING database using a high-confidence interaction score threshold (> 0.9). After exclusion of 29 disconnected proteins, the remaining 56 genes formed a network comprising 170 interaction pairs, with an average node degree of 4.15 and a local clustering coefficient of 0.449 (enrichment $P < 1.0 \times 10^{-16}$) (**Figure 3C**). The top 20 candidate genes (*FOS*, *NR4A1*, *NFKBIA*, *DUSP1*, *CDKN1A*, *FOSL2*, *TNFAIP3*, *EGR2*, *GADD45B*, *BCL3*, *TGFB1*, *DUSP5*, *MAPK7*, *ZFP36*, *MMP9*, *CREM*, *MAP2K7*, *DUSP2*, *MAP3K8*, and *NFIL3*) were subsequently selected from the 56 candidate genes based on Degree scores.

Early Growth Response Protein 2 (EGR2) and Transforming Growth Factor-β1 (TGFB1): Potential Biomarker Genes in AR

Among the 3 classifiers, the GLM and RF models achieved high AUC values in the training dataset (GSE75011; internal evaluation only) (**Figure 4A-4C, Supplementary Table 3**). The RF

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]

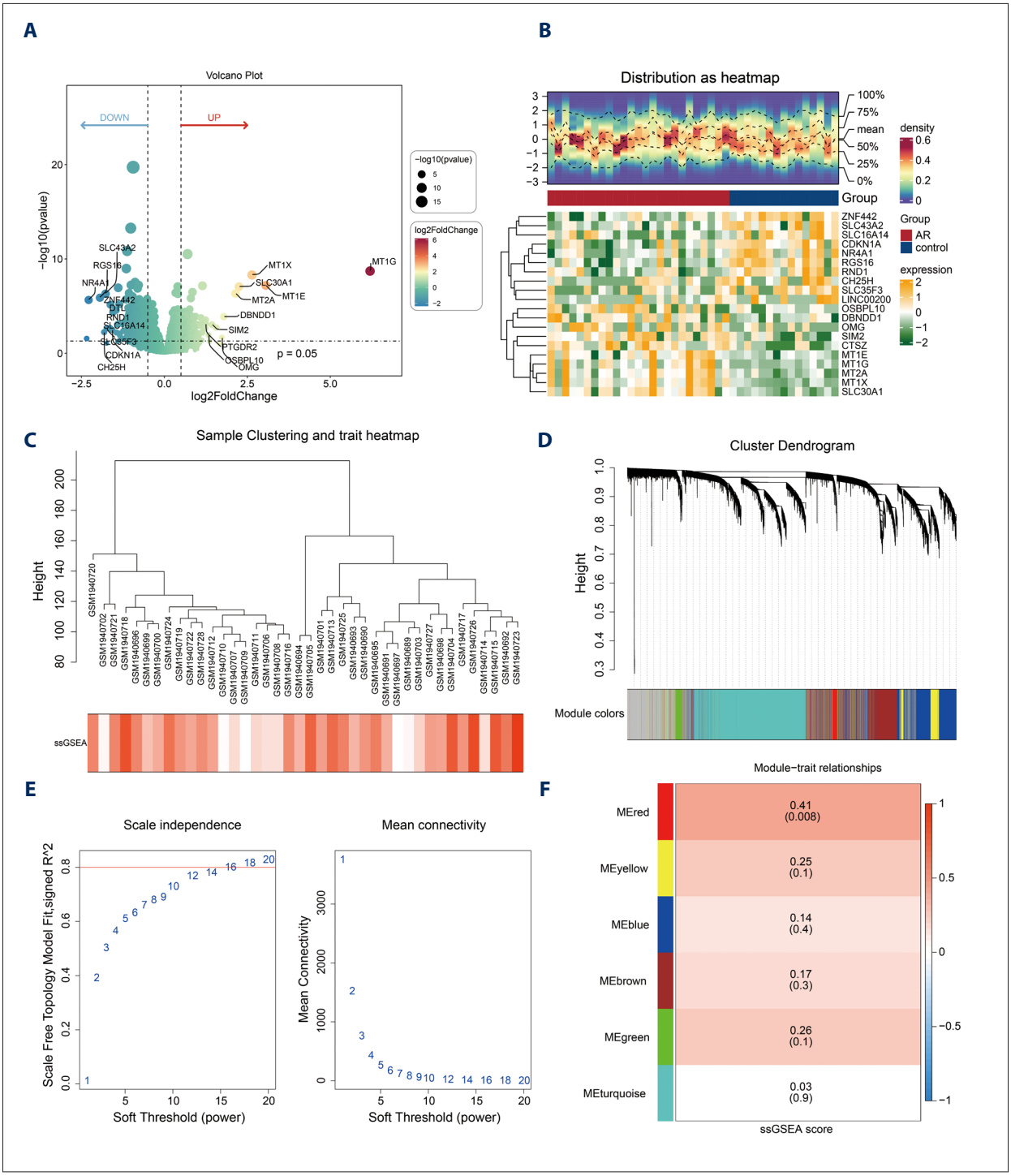


Figure 1. Differential allergic rhinitis (AR)-related gene expression and analysis. (**A**) Volcano plot of differentially expressed genes in AR samples. (**B**) Heat maps and density plots of the top 10 upregulated and downregulated genes in AR samples. (**C**) Cluster analysis of the GSE75011 dataset. (**D**) Determination of soft-thresholding power in weighted gene co-expression network analysis. **Left panel** illustrates the relationship between the scale-free topology fit index (signed R^2) and soft-thresholding power, with the red horizontal line indicating the threshold of $R^2 = 0.8$. **Right panel** shows average connectivity versus soft-thresholding power, demonstrating that average connectivity approaches zero as soft-thresholding power increases. (**E**) Clustering dendrogram of gene modules. (**F**) Relationships between gene modules and chemokine-related gene (CRG) scores. Abbreviation: ssGSEA, single-sample gene set enrichment analysis

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]

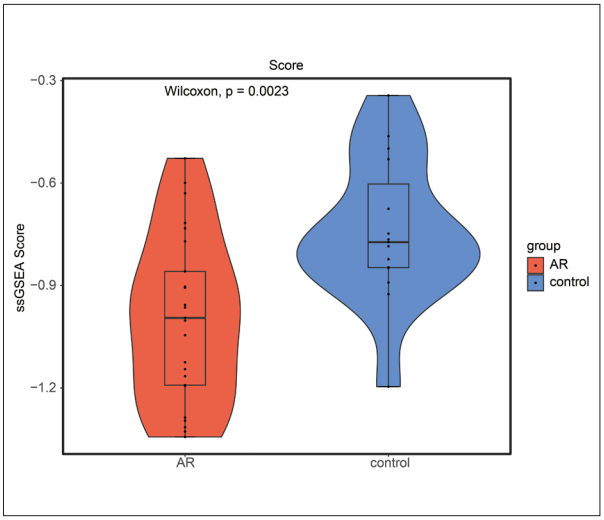


Figure 2. Distribution of single-sample gene set enrichment analysis (ssGSEA) scores for chemokine-related genes in allergic rhinitis and control groups. Box-and-line plots showing the distribution of ssGSEA scores for chemokine-related genes in the AR group (orange) and control group (blue).

model showed the lowest RMSR and RMSE in this internal assessment; however, no independent external validation was available to confirm model generalizability. The confusion matrix further indicated that the RF model effectively distinguished AR samples from control samples within the training dataset (**Figure 5A-5C**). Additionally, the RF model demonstrated the highest accuracy, precision, recall, and F1 score among the 3 classifiers (**Figure 5D-5F, Supplementary Table 3**). Based on these internally derived performance metrics, the RF model was considered the optimal classifier in this study. However, because all metrics were calculated using the same dataset used for model training and were not validated using an independent external dataset, the reported performance estimates may be subject to optimistic bias and should be interpreted with caution. The RF model identified optimal features by calculating the Gini index for each of the 20 candidate genes. The top 5 genes with the highest Gini index were selected as candidate biomarkers: *DUSP5*, *NFKBIA*, *EGR2*, *NR4A1*, and *TGFB1* (**Figure 4D, Supplementary Table 4**). Subsequent expression analysis identified *EGR2* and *TGFB1* as potential biomarkers for AR because both genes exhibited significantly downregulated expression in AR samples compared with control samples across the GSE75011 and GSE50223 datasets ($P < 0.05$) (**Figure 4E, 4F**). We note that the machine learning analysis in this study did not incorporate cross-validation or an independent training-test split strategy, which may have increased the risk of overfitting. Furthermore, all statistical analyses were based on unadjusted raw *P* values without correction for multiple testing. Thus, *EGR2* and *TGFB1* should be considered exploratory candidate biomarkers, rather than definitive diagnostic biomarkers.

Exploratory Construction and Internal Evaluation of a Nomogram Based on Potential Biomarkers

A nomogram was developed to explore the potential utility of the identified biomarkers for estimating the probability of AR within the discovery dataset. As shown in **Figure 6A**, higher total scores corresponded to a greater predicted probability of AR. The Hosmer-Lemeshow test yielded a *P* value of 0.141 (exceeding 0.05), and the calibration curve slope was close to 1, indicating acceptable internal consistency of the nomogram in this exploratory setting (**Figure 6B**). ROC analysis yielded an AUC of 0.775 during internal exploratory evaluation (**Figure 6C**). However, all performance metrics were solely derived from internal analyses without independent validation and therefore represent exploratory predictive potential, rather than validated clinical performance.

GSEA of Biomarkers

GSEA was performed to investigate biological functions associated with the biomarkers. *EGR2* and *TGFB1* were enriched in 6 and 25 pathways, respectively ($P < 0.05$). Shared enriched pathways included the MAPK signaling pathway and the p53 signaling pathway. Additionally, *EGR2* was associated with pathways such as circadian rhythm-mammal, spliceosome, and tryptophan metabolism (**Figure 7A**). *TGFB1* was also enriched in the chemokine signaling pathway, neurotrophin signaling pathway, and propanoate metabolism (**Figure 7B**). *TGFB1* participation in chemokine-related signaling pathways may directly influence immune cell recruitment, whereas *EGR2*-associated pathways might indirectly regulate chemokine expression patterns. Taken together, our findings support the relevance of these chemokine-related biomarkers in AR.

M2 Macrophages and Activated Memory CD4+ T Cells Exhibited Reduced Infiltration in AR

The estimated proportions of 22 immune cell types in samples from the GSE75011 dataset are shown in **Figure 7C**. All samples were retained after filtering (n = 40). Five DICs were identified between AR and control groups, including activated dendritic cells, M2 macrophages, activated natural killer cells, activated memory CD4+ T cells, and naïve CD4+ T cells ($P < 0.05$). Activated dendritic cells, M2 macrophages, and activated memory CD4+ T cells exhibited lower inferred relative proportions in AR samples, whereas activated natural killer cells and naïve CD4+ T cells showed higher inferred relative proportions in AR samples ($P < 0.05$) (**Figure 7D, Supplementary Table 5**). Spearman correlation analysis using thresholds of $|cor| > 0.3$ and $P < 0.05$ demonstrated that *TGFB1* was significantly negatively correlated with activated memory CD4+ T cells ($cor = -0.64$, $P < 0.05$) and positively correlated with M2 macrophages ($cor = 0.57$, $P < 0.05$) (**Figure 7E**). Such associations suggest that *TGFB1* can influence immune imbalance in AR by

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]

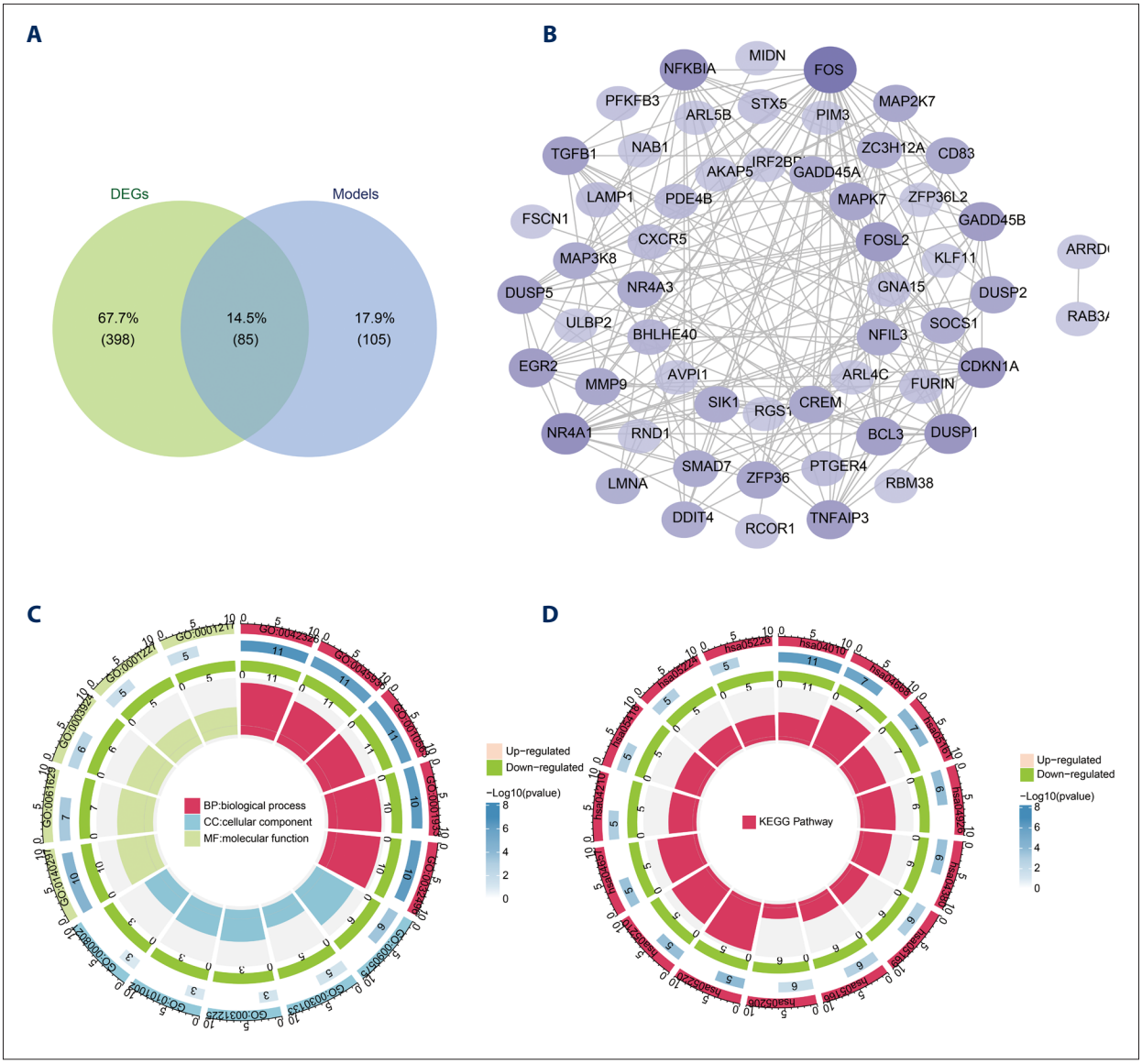


Figure 3. Identification and enrichment analysis of candidate genes and protein-protein interaction (PPI) networks. (**A**) Venn diagram of candidate genes. (**B**) Gene Ontology and Kyoto Encyclopedia of Genes and Genomes pathway analyses. (**C**) PPI network of candidate genes. Nodes represent genes, and edges indicate interactions between genes. Abbreviation: DEGs, differentially expressed genes.

regulating chemokine-mediated recruitment of these immune cell populations, further supporting a role for chemokines in AR-associated immune dysregulation. We note that the immune cell fractions reported in this study were inferred from bulk transcriptomic data using the CIBERSORT deconvolution algorithm. These values represent estimated relative proportions of immune cell types within samples, rather than absolute cell counts; they also reflect computational predictions, rather than direct experimental measurements.

Tretinoin Might Represent a Potential Therapeutic Candidate for AR

Potential therapeutic agents targeting *EGR2* or *TGFB1* were identified using the DGIdb database. For *TGFB1*, 58 drugs with available interaction scores were retrieved. These drugs were ranked in descending order according to DGIdb interaction scores, which serve as qualitative rather than quantitative indicators of interaction strength. The top 3 drugs—bovine cartilage, idoxuridine, and tretinoin—were selected for subsequent molecular docking analysis. For *EGR2*, only 1 candidate drug, tretinoin, was identified (**Figure 8A**). To investigate

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]

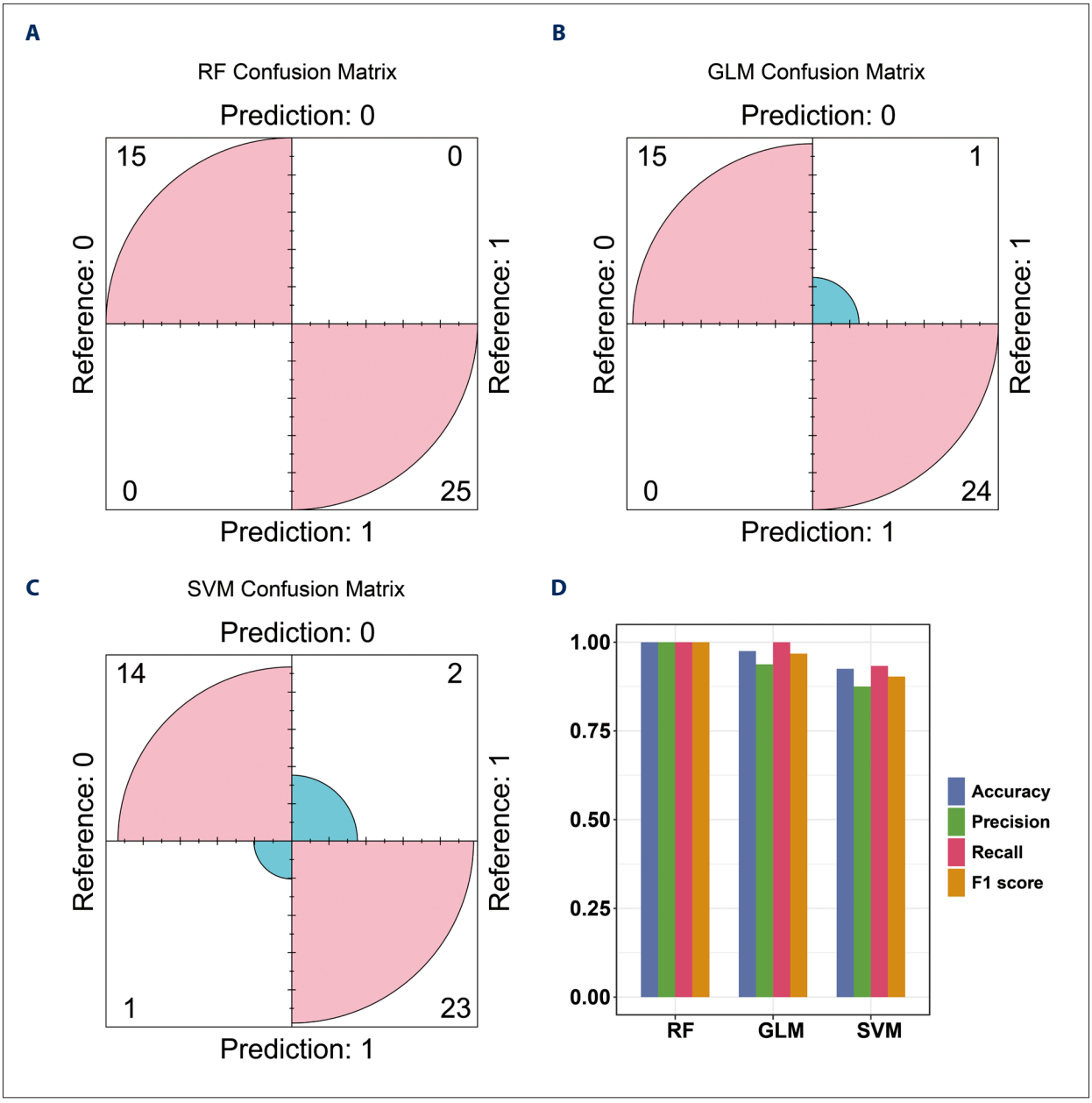


Figure 4. Identification and validation of potential biomarkers in AR. (**A**) Box-and-line plots showing the distributions of absolute residuals for the random forest (RF; light blue), generalized linear model (GLM; pink), and support vector machine (SVM; light green) models. (**B**) Receiver operating characteristic (ROC) curves for different models. (**C**) Feature importance ranking and root mean square error (RMSE) loss across different models. (**D**) Candidate gene feature importance based on the RF classifier. (**E**) Gene expression levels in the training dataset ($P < 0.05$). (**F**) Gene expression levels in the validation dataset ($P < 0.05$).

binding affinity between the predicted drugs and the 2 potential biomarkers, molecular docking analysis was performed. The binding energies between *TGFB1* (PDB ID: 4KV5) and bovine cartilage (PubChem CID: 24766), idoxuridine (PubChem CID: 5905), and tretinoin (PubChem CID: 444795) were -7.2 kcal/mol, -6.2 kcal/mol, and -7.5 kcal/mol, respectively—all below the threshold of -5 kcal/mol, indicating favorable binding affinity (**Figure 8B-8D, Supplementary Table 6**). Molecular docking analysis for *EGR2* could not be performed because its 3-dimensional structure was unavailable in the PDB database. We emphasize that the drug prediction and molecular docking results presented in this study are preliminary exploratory findings intended to generate potential therapeutic hypotheses. The efficacy, safety, and mechanisms of action of tretinoin as

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]

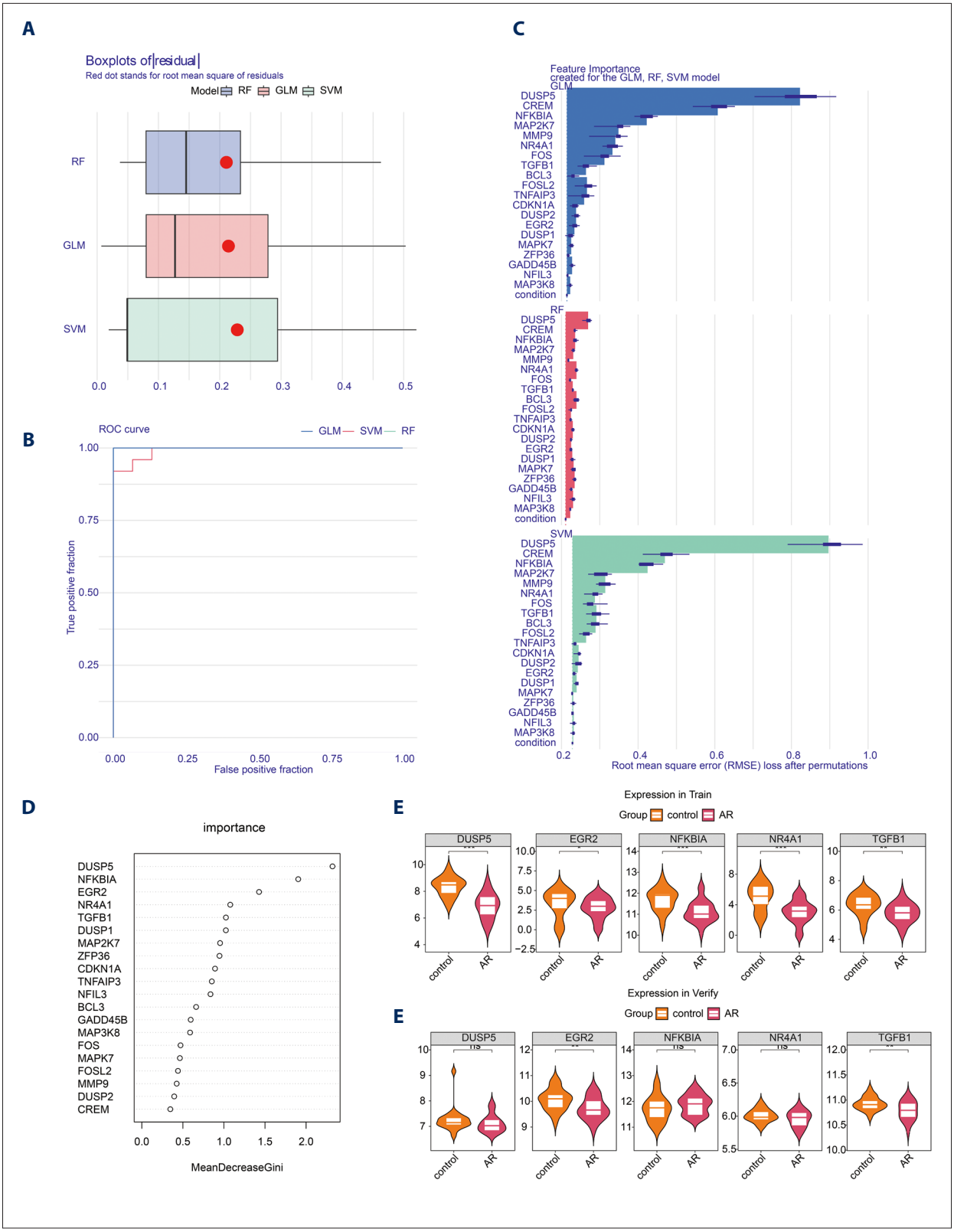


Figure 5. Further validation of random forest (RF) model reliability. (**A-C**) Confusion matrices for the RF, generalized linear model (GLM), and support vector machine (SVM) models. (**D-F**) Accuracy, precision, recall, and F1 scores of the GLM, SVM, and RF models.

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]

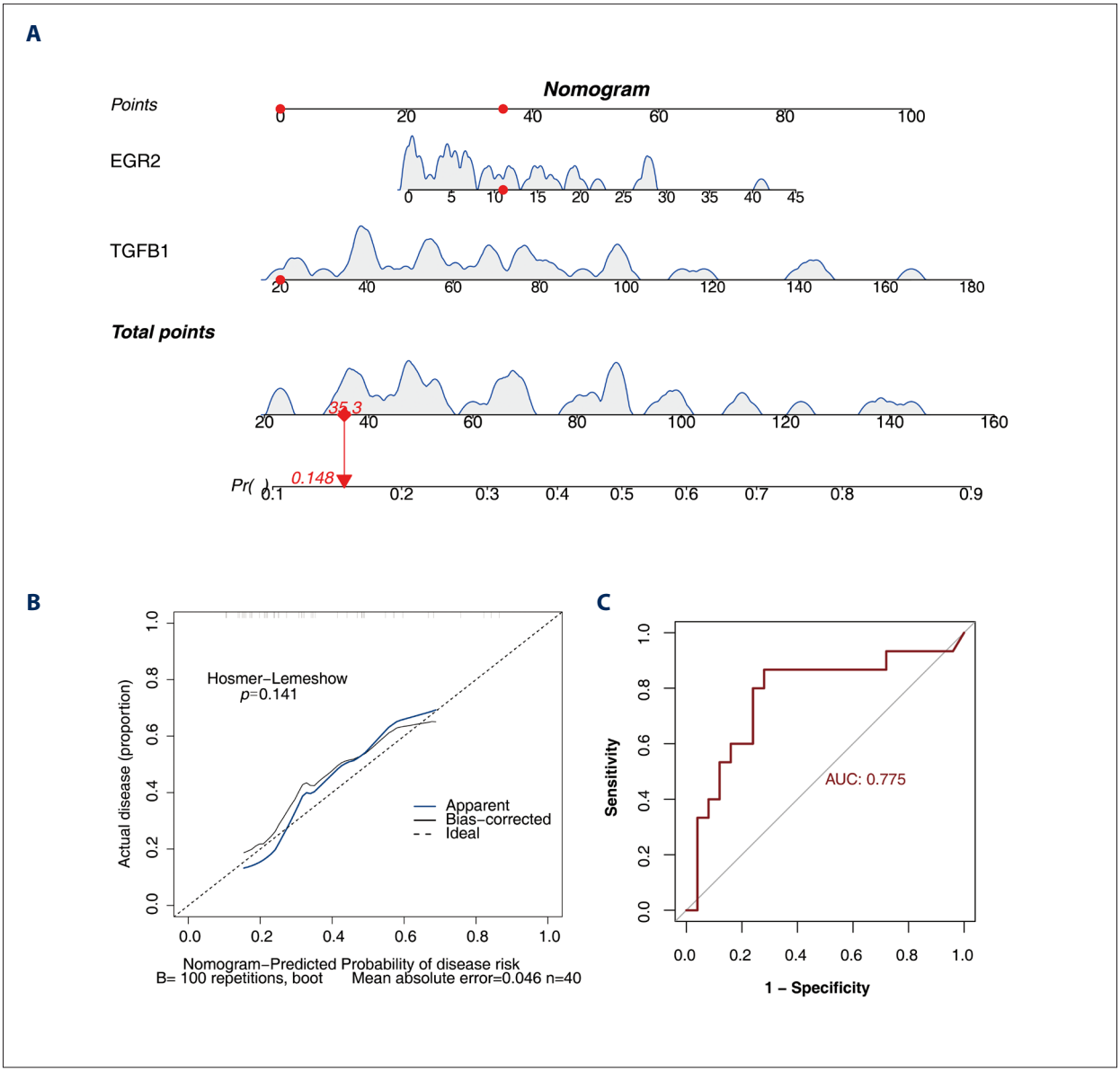


Figure 6. Probability prediction model for AR. (**A**) Nomogram constructed to predict AR probability based on *EGR2* and *TGFB1*. (**B**) Calibration curve of the nomogram for predicting disease probability (Hosmer-Lemeshow $P = 0.141$). (**C**) Receiver operating characteristic (ROC) analysis of the nomogram (area under the curve = 0.775).

a candidate drug targeting both *EGR2* and *TGFB1* in AR require systematic validation through future experimental studies.

EGR2 and TGFB1 Were Downregulated in AR Samples

Consistent with findings from the GSE75011 and GSE50223 datasets, RT-qPCR validation using ex vivo whole blood samples (5 patients with AR vs 5 healthy controls; Third People's Hospital of Hefei) demonstrated significantly reduced expression levels of *EGR2* ($P = 0.0213$) and *TGFB1* ($P = 0.0196$) in AR samples (**Figure 9A, 9B, Supplementary Table 7**).

Discussion

AR affects approximately 10% to 30% of the global population; thus, optimization of its diagnostic and therapeutic approaches remains a major research focus [3]. Although pharmacological treatments (eg, antihistamines and corticosteroids) and allergen immunotherapy have been widely applied, conventional pharmacotherapy is often associated with side effects and limited long-term efficacy, whereas allergen immunotherapy is constrained by prolonged treatment duration and high costs [33]. In recent years, numerous bioinformatics studies based on high-throughput data have sought to identify key

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]

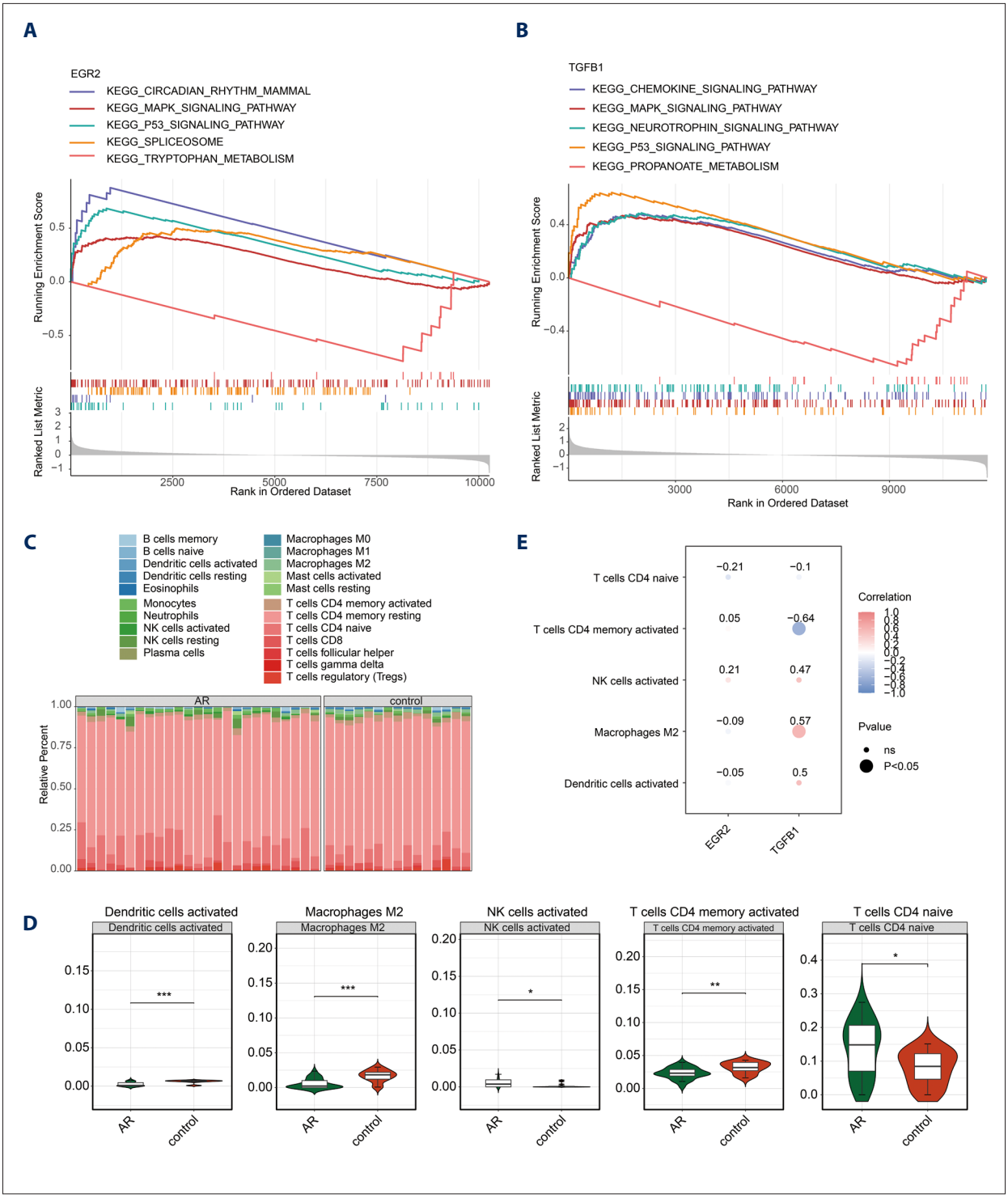


Figure 7. Gene set enrichment analysis (GSEA) and immune infiltration analysis of biomarkers. (**A, B**) GSEA results for *EGR2* (**A**) and *TGFB1* (**B**). (**C**) Estimated proportions of 22 immune cell types in samples from the GSE75011 dataset. (**D**) Distribution of 5 differentially infiltrated immune cells (DICs) between AR and control groups. (**E**) Correlations between biomarkers and immune cell populations.

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]

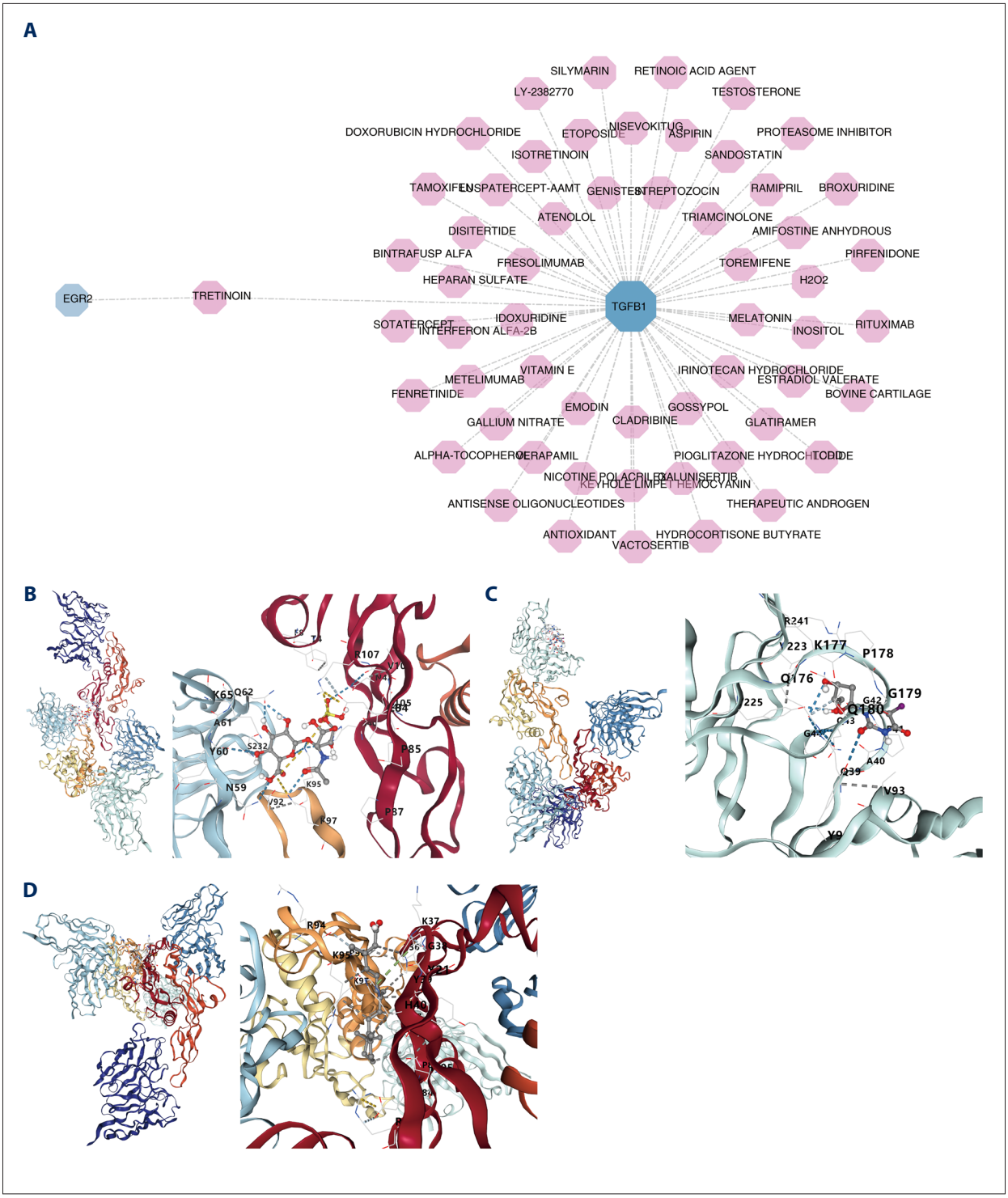


Figure 8. Biomarker-based drug prediction and molecular docking analysis. (**A**) Interaction network of *TGFB1* and associated drugs. (**B**) Molecular docking analysis of *TGFB1* with bovine cartilage. (**C**) Molecular docking analysis of *TGFB1* with idoxuridine. (**D**) Molecular docking analysis of *TGFB1* with tretinoin.

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]

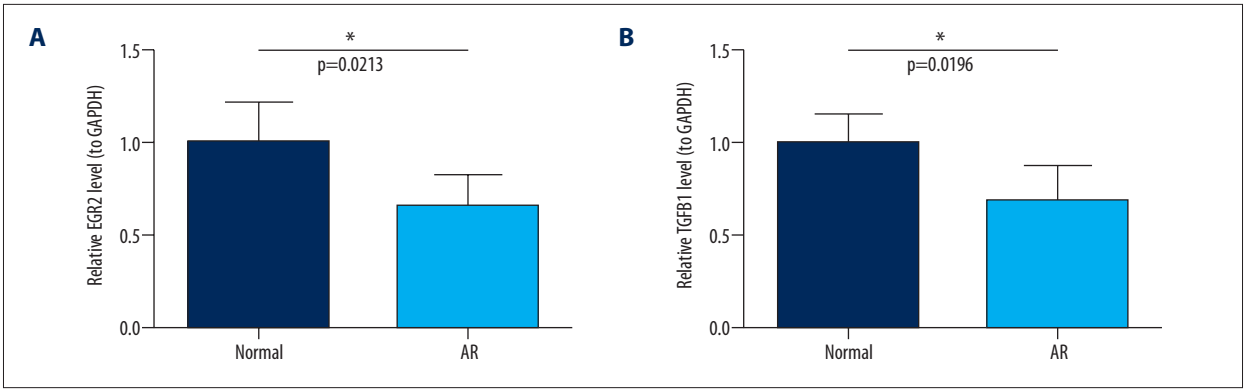


Figure 9. Validation of differential biomarker mRNA expression in AR and control samples by reverse transcription quantitative polymerase chain reaction (RT-qPCR). Differential mRNA expression levels of biomarkers were validated in AR samples (n = 5) and control samples (n = 5) ($P < 0.05$). (**A**) *EGR2*. (**B**) *TGFB1*.

molecular markers associated with AR. For example, Zhu et al identified *INPP5D* and *KIF16B* as potential key genes in AR by integrating multi-omics and single-cell RNA sequencing data; their work highlighted the importance of Th1 cells [34]. Li et al used single-cell transcriptomic analysis to explore the regulatory effects of Peiyuan Tong-qiao decoction on immune cell subsets in AR [35]. Unlike these previous studies, the present study used CRGs as the starting point and implemented WGCNA combined with 3 machine learning algorithms (GLM, RF, and SVM) for stepwise screening, ultimately identifying 2 immune regulatory genes, *EGR2* and *TGFB1*. Whereas previous studies primarily focused on directly identifying the genes with greatest differential expression, the present study emphasized gene module membership within co-expression networks and the potential role of immune regulatory axes, highlighting the complementary nature of this screening strategy.

Although our CRG-based bioinformatics screening identified *EGR2* and *TGFB1* as potential biomarkers for AR, subsequent GSEA and immune infiltration analyses suggested that their principal mechanistic relevance lies in broader immunoregulatory functions rather than direct chemokine activity. GSEA indicated that both genes are involved in key signaling pathways, including MAPK and p53, which regulate immune cell activation, proliferation, apoptosis, and inflammatory responses [36,37]. Immune infiltration analysis revealed significant associations between *TGFB1* expression and both M2 macrophages and activated memory CD4+ T cells, supporting their roles as immune modulators within the inflammatory microenvironment of AR. Collectively, these findings suggest that *EGR2* and *TGFB1* primarily function as immunoregulatory biomarkers that may contribute to AR pathogenesis via modulation of immune cell function and inflammatory processes.

EGR2 is a zinc finger transcription factor belonging to the early growth response (Egr) gene family [38]. It plays a pivotal role in T-cell activation, tolerance, and function [39]. *EGR2* is also essential for maintaining immune homeostasis by promoting adaptive immune responses and regulating inflammation [40]. Under inflammatory conditions in vitro, internal ribosome entry site (IRES)-dependent translation of *EGR2* is rapidly induced, highlighting its role as an immediate responder during inflammation [41]. Moreover, *EGR2* has been implicated in various inflammatory diseases. For instance, in autoimmune uveitis, *EGR2* alleviates disease severity by modulating retinal microglial phenotypes through activation of GDF15 [42]. In systemic lupus erythematosus, *EGR2* gene polymorphisms have been associated with disease susceptibility [43]. These findings suggest that *EGR2* may similarly influence AR progression through related mechanisms. Furthermore, AR is characterized by an imbalance in Th1/Th2 cell-mediated immune responses [8]. *EGR2* may regulate the severity of allergic reactions in patients with AR by influencing the differentiation and function of Th1 and Th2 cells. In addition to transcription factors (eg, *EGR2*) and cytokines (eg, TGF-β1), certain bioactive compounds such as chlorogenic acid have been reported to alleviate AR symptoms via modulation of Th1/Th2 balance [44]. Overall, *EGR2* may contribute to the occurrence and progression of AR by regulating immune homeostasis, inflammatory responses, and immune cell function, thus providing valuable insights into AR mechanisms and potential therapeutic interventions.

The *TGFB1* gene encodes TGF-β1, a multifunctional cytokine with critical roles in various cellular processes, including regulation of immune responses [45]. TGF-β1 maintains immune tolerance primarily by inducing differentiation of regulatory T cells. Downregulation of TGF-β1 can reduce regulatory T cell populations, weakening suppression of Th2 responses and promoting release of proinflammatory cytokines such as IL-4, IL-5, and IL-13; these processes exacerbate eosinophilic infiltration and nasal mucosal hyperresponsiveness [46-48]. In conjunction with IL-6, TGF-β1 can drive Th17 cell differentiation, enhancing neutrophil recruitment and contributing to tissue remodeling during the chronic inflammatory phase of AR [49].

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]

Immune infiltration analysis showed that *TGFB1* expression was significantly positively correlated with M2 macrophages (cor = 0.57) and negatively correlated with activated memory CD4+ T cells (cor = -0.64). These findings suggest that TGF-β1 can promote M2 macrophage polarization through activation of Smad-independent MAPK pathways. Reduced TGF-β1 expression may impair the anti-inflammatory and tissue repair functions mediated by M2 macrophages, thus contributing to AR chronicity [50,51]. Additionally, low TGF-β1 levels might diminish its inhibitory effects on memory CD4+ T cells, potentially accelerating Th2 effector reactivation and triggering recurrent allergic symptoms. Overall, *TGFB1* may profoundly influence the onset and progression of AR through regulation of immune cell differentiation and function; its dysregulation offers a promising target for AR diagnosis and treatment.

GSEA in this study revealed that the selected biomarkers, *EGR2* and *TGFB1*, were significantly enriched in the MAPK and p53 signaling pathways. This finding suggests that their roles in AR are closely associated with these 2 core pathways. Given our results and existing literature, we propose the following mechanistic hypothesis. In AR, allergens may activate immune cell surface receptors (eg, IgE receptors), thereby triggering a MAPK signaling cascade. Activation of the MAPK pathway, including p38 MAPK and Janus kinase, can promote the release of proinflammatory cytokines and drive nasal mucosal inflammation [52,53]. As a known downstream target of MAPK signaling, downregulated *EGR2* expression may contribute to inflammatory regulation in AR by altering feedback regulation within this pathway [41,54,55]. Similarly, *TGFB1* may interact with the MAPK pathway; its downregulation may impair regulation of epithelial and immune cell functions, including regulatory T cell and Th17 cell activity, thus contributing to immune dysregulation and chronic tissue remodeling in AR [49,56,57]. Additionally, the p53 pathway plays a critical role in regulating apoptosis, inflammation, and maintenance of epithelial barrier integrity [58-60]. Enrichment of *EGR2* and *TGFB1* in the p53 pathway, as identified in the present study, suggests their involvement in regulating nasal epithelial barrier function in AR via modulation of p53-mediated cellular homeostasis. Previous studies have shown that *EGR2* can act as a transcriptional target of p53 and participate in p53-mediated apoptosis [61], whereas TGF-β1 may cooperate with p53 to maintain epithelial integrity [62]. Therefore, we hypothesize that downregulation of *EGR2* and *TGFB1* contributes to persistent inflammation and epithelial barrier dysfunction in AR by disrupting the normal function of the MAPK/p53 signaling network. Collectively, the present work links reduced *EGR2* and *TGFB1* expression with the potential involvement of MAPK/p53 signaling in AR through GSEA findings, providing an important basis and theoretical framework for future mechanistic investigations. Specific molecular interactions underlying these pathways require further experimental validation.

Immune infiltration analysis in this study revealed significant differences in the inferred relative proportions of activated memory CD4+ T cells and M2 macrophages in the AR group. *TGFB1* expression was negatively correlated with activated memory CD4+ T cells and positively correlated with M2 macrophages. These findings suggest that *TGFB1* participates in regulating the balance between these 2 key immune cell populations within the local immune microenvironment of AR. Activated memory CD4+ T cells, which play central roles in adaptive immunity, may predominantly polarize toward a Th2 phenotype in AR and thus promote allergic inflammatory responses such as eosinophil infiltration [63,64]. In contrast, M2 macrophages possess anti-inflammatory and tissue repair functions, although their profibrotic properties may also contribute to chronic AR progression [50,51]. Notably, GSEA results demonstrated that both *TGFB1* and *EGR2* were significantly enriched in the MAPK and p53 signaling pathways. Based on these associations, we hypothesize that downregulation of *TGFB1* impairs its ability to induce M2 macrophage polarization through altered MAPK pathway activity [65], while also reducing its inhibitory effects on memory CD4+ T cells. In parallel, dysregulation of the MAPK/p53 signaling network may collectively promote excessive Th2 responses and macrophage dysfunction, thereby exacerbating inflammation and tissue remodeling in AR. Previous research has also shown that *EGR2* plays important roles in regulating T-cell function and macrophage programming [66]; its downregulation may further disrupt immune homeostasis. In summary, the present study identified associations between *TGFB1* and specific immune cell phenotypes, suggesting that the underlying mechanisms involve MAPK/p53 signaling. Our findings support the hypothesis that targeting *TGFB1*/*EGR2* via modulation of the MAPK/p53 pathway to restore immune balance between memory T cells and macrophages offers a potential strategy for intervening in AR progression. However, these proposed mechanisms require further experimental validation.

Through database-based prediction, the present study identified tretinoin as a compound capable of simultaneously targeting *EGR2* and *TGFB1*, suggesting its capacity to exert multiple therapeutic effects in AR via regulation of these 2 genes. Molecular docking analysis demonstrated high binding affinity between tretinoin and *TGFB1*, indicating promising targeting potential. From a clinical translation perspective, tretinoin, a drug already used for various inflammatory skin diseases [67], offers a novel therapeutic strategy for AR through its immunomodulatory and antifibrotic properties. Particularly in patients with chronic AR accompanied by tissue remodeling, tretinoin may alleviate symptoms and delay disease progression by inhibiting aberrant activation of the MAPK/p53 pathway, regulating macrophage polarization, and modulating memory T-cell responses. However, the drug prediction and molecular docking findings in this study are based on computational simulations and have not yet been

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]

experimentally validated. Future studies should utilize in vitro and in vivo models to further investigate the effects of tretinoin on *EGR2*/*TGFB1* expression and downstream immune cell functions, as well as to evaluate its efficacy and safety.

The identification of *EGR2* and *TGFB1* as potential immunomodulatory biomarkers provides new diagnostic and therapeutic avenues for AR research. Current AR diagnosis primarily relies on clinical symptoms and a limited number of biomarkers, with a lack of highly specific indicators [4]. In the present study, downregulated expression levels of *EGR2* and *TGFB1* were consistently observed across both public datasets and clinical samples, suggesting their utility as auxiliary diagnostic biomarkers for improving early AR detection. Additionally, current therapies mainly focus on symptom relief and often exhibit limited long-term efficacy with associated side effects [68]. Our findings indicate that *EGR2* and *TGFB1* may regulate immune cell infiltration and inflammatory responses in AR through key signaling pathways such as MAPK and p53, providing a theoretical basis for the development of novel targeted immunomodulatory therapies. Notably, drug prediction analysis suggested that tretinoin can simultaneously target both genes; molecular docking demonstrated high affinity between tretinoin and *TGFB1*, offering new perspectives for combined therapeutic strategies in AR. Overall, *EGR2* and *TGFB1* not only broaden our understanding of the immunoregulatory network underlying AR but also demonstrate translational potential for improving diagnostic precision and advancing targeted therapeutic approaches.

Nevertheless, this study has several limitations. First, no corrections for multiple testing were applied in any statistical analyses; the machine learning models and nomogram were constructed and evaluated using the same dataset without independent external validation or cross-validation strategies, which may increase the risk of overfitting. Thus, the identified biomarkers should be considered exploratory, hypothesis-generating findings rather than validated diagnostic markers. Second, the RT-qPCR validation cohort was very small (n = 5 per group), no formal power calculation was performed, and no protein-level validation was conducted. Accordingly, these findings should be regarded as preliminary supportive evidence only. To address these limitations, future studies will aim to expand clinical sample sizes and conduct multicenter validation with pre-specified power calculations. Independent training-test split strategies and external validation cohorts will be incorporated to improve model robustness. Corrections for multiple testing will also be applied to control the false-positive rate. Additionally, RT-qPCR validation cohorts will be expanded, and protein-level validation using techniques such as Western blotting or immunohistochemistry will be performed. Finally, in vitro and in vivo studies are planned to clarify the functional roles of *EGR2* and *TGFB1* in AR pathogenesis, then facilitate translation of these exploratory findings into clinical applications.

Conclusions

Through CRG-based bioinformatics screening and preliminary experimental validation, this study identified *EGR2* and *TGFB1* as potential immunomodulatory biomarkers associated with AR. Our nomogram based on *EGR2* and *TGFB1* may serve as an exploratory analytical tool. Although the findings are hypothesis-generating in nature, they suggest that reduced expression levels of *EGR2* and *TGFB1* are associated with AR pathogenesis and can offer utility as auxiliary diagnostic biomarkers. Additionally, these biomarkers may represent potential targets for future therapeutic interventions, as suggested by predictive drug analyses identifying compounds such as tretinoin. However, our conclusions require further validation through larger clinical studies, functional experiments, and mechanistic investigations. Overall, this study provides new theoretical insights and research directions for precision diagnosis and targeted therapy in AR.

Acknowledgments

We thank our collaborators at Anhui Medical University for their contributions to the experimental work. Finally, we thank our families for their understanding and support throughout the research and writing process.

Author Statement

All figures presented in this study were generated by the authors using R software, Cytoscape, CB-Dock, and PyMOL based on publicly available datasets and the authors' own experimental data. No figures were reproduced from other sources.

Department and Institution Where Work Was Done

Department of Otolaryngology, Head and Neck Surgery, First Affiliated Hospital of Anhui Medical University, Hefei, Anhui, PR China.

Patient Consent

Written informed consent was obtained from all participants prior to enrollment in this study.

Declaration of Figures' Authenticity

All figures submitted have been created by the authors who confirm that the images are original with no duplication and have not been previously published in whole or in part.

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]

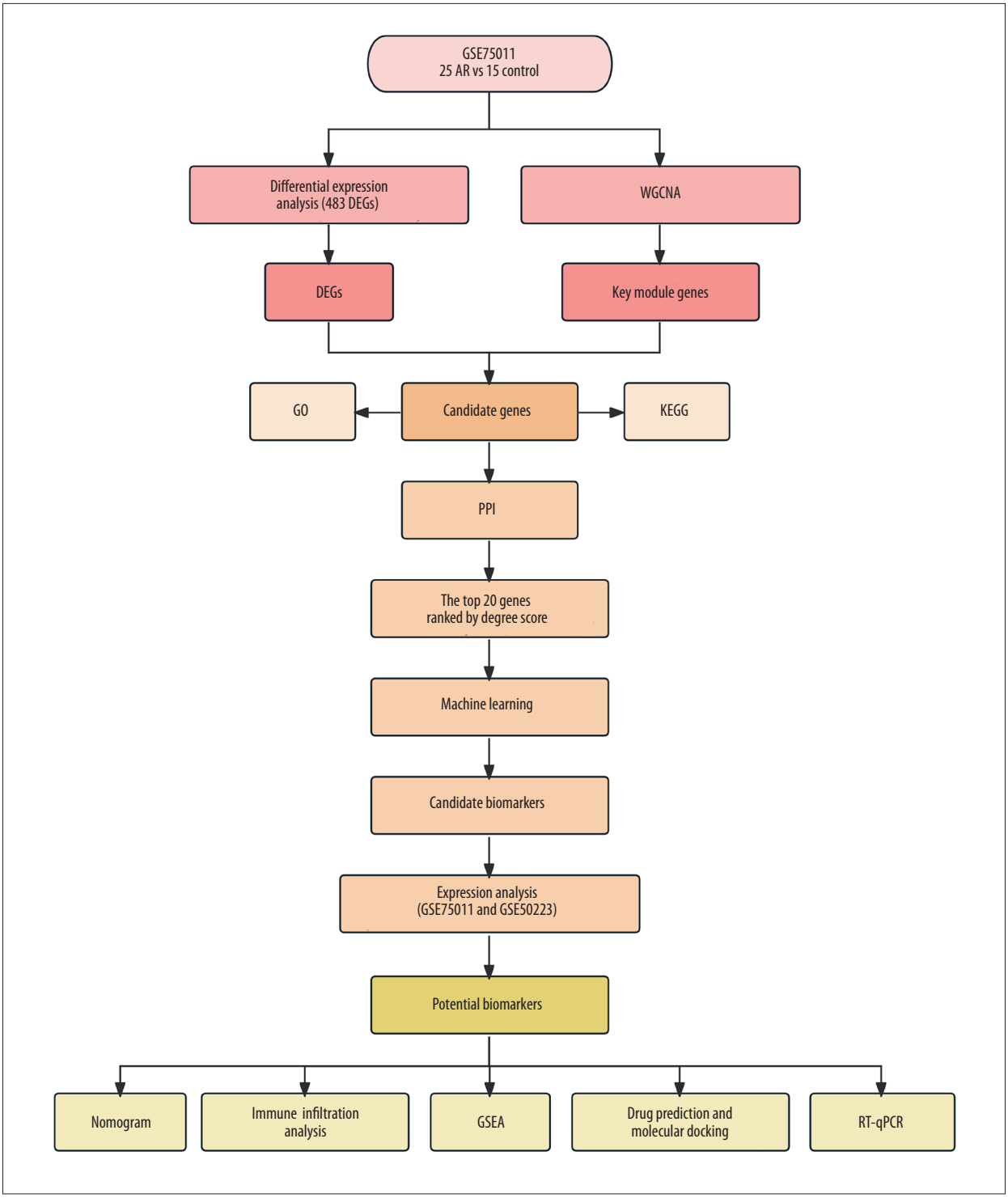
Supplementary Materials

Supplementary Table 1. Clinical characteristics of the validation cohort.

Number	Gender	Age (years)	Duration of disease (months)	VAS score	Allergen	SPT level	Medication history prior to skin test
AR patient 1	Female	32	60	7	*Dermatophagoides pteronyssinus*	Grade II	Discontinuation of oral antihistamines for 2 weeks and intranasal corticosteroids for 4 weeks
					Dermatophagoides farinae	Grade II	
					Cladosporium cladosporioides	Grade III	
AR patient 2	Male	41	24	8	*Dermatophagoides pteronyssinus*	Grade II	Discontinuation of oral antihistamines for 2 weeks
					Dermatophagoides farinae	Grade II	
AR patient 3	Female	35	36	6	*Dermatophagoides pteronyssinus*	Grade III	Discontinuation of intranasal corticosteroids for 4 weeks and leukotriene receptor antagonists for 2 weeks
					Dermatophagoides farinae	Grade III	
					Cladosporium cladosporioides	Grade III	
AR patient 4	Female	26	84	9	*Dermatophagoides pteronyssinus*	Grade II	Discontinuation of oral antihistamines for 2 weeks
					Dermatophagoides farinae	Grade II	
					Cladosporium cladosporioides	Grade III	
AR patient 5	Male	24	15	5	Cat hair	Grade II	Discontinuation of intranasal corticosteroids for 4 weeks
Control 1	Male	40	–	0	–	0	–
Control 2	Female	42	–	0	–	0	–
Control 3	Female	35	–	0	–	0	–
Control 4	Male	27	–	0	–	0	–
Control 5	Female	25	–	0	–	0	–

Note: During the collection of clinical samples, all participants (including the AR group and the control group) were excluded if they had the following conditions: (1) complicated with asthma, atopic dermatitis, or other systemic immune disorders;(2) suffering from chronic rhinosinusitis, nasal polyps, or other structural nasal diseases; (3) having a history of acute respiratory tract infection within 4 weeks prior to enrollment; (4) having a history of nasal surgery or undergoing allergen-specific immunotherapy; (5) being pregnant or lactating. The diagnosis of AR patients was strictly in accordance with the criteria of the ARIA (2020) guideline, and all of the following conditions were required to be satisfied: (1) presenting with at least two typical symptoms among nasal obstruction, rhinorrhea, sneezing, and nasal pruritus, with symptom duration exceeding 4 weeks;(2) exhibiting a positive reaction to at least one common inhaled allergen (*Dermatophagoides pteronyssinus*, *Dermatophagoides farinae*, *Cladosporium cladosporioides*, cat hair) by SPT. Before SPT administration, intranasal or oral glucocorticoids were discontinued for at least 4 weeks, and antihistamines (including both intranasal sprays and oral formulations) were discontinued for at least 2 weeks. The diagnostic criteria for SPT were as follows: wheal diameter = 1/2 × (longest diameter + shortest diameter); skin index (SI) = diameter of allergen-induced wheal/diameter of positive control wheal. Grade 0 (SI < 0.3): negative; Grade I (0.3 < SI < 0.5): +; Grade II (0.5 ≤ SI < 1.0): ++; Grade III (1.0 ≤ SI < 2.0): +++; Grade IV (SI ≥ 2.0): ++++. Healthy controls had no history of AR or asthma, and tested negative for all evaluated allergens in SPT.
AR: allergic rhinitis; VAS: Visual analogue scale; SPT: skin prick test.

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]



Supplementary Figure 1. Study workflow diagram.

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]

Supplementary Table 2. Sample distribution across each analytical step.

Analysis step	Dataset	Initial samples (AR/Control)	Excluded samples (Reason)	Final samples (AR/Control)
Data preprocessing	GSE75011	25/15	0	25/15
DEG analysis	GSE75011	25/15	0	25/15
WGCNA	GSE75011	25/15	0	25/15
Machine learning	GSE75011	25/15	0	25/15
CIBERSORT	GSE75011	25/15	0 (all p<0.05)	25/15
Correlation analysis	GSE75011	25/15	0	25/15
External validation	GSE50223	21/21	0	21/21

Supplementary Table 3. Performance metrics for the 3 machine learning models (GLM, SVM, RF).

Model	RMSR	RMSE	AUC	Accuracy	Precision	Recall	F1 score
GLM	0.221	1.001	1.000	0.975	0.938	1.000	0.968
SVM	0.229	0.502	0.992	0.925	0.875	0.933	0.903
RF	0.211	0.467	1.000	1.000	1.000	1.000	1.000

Supplementary Table 4. Gini indices of candidate biomarkers.

Rank	Gene	MeanDecreaseGini
1	*DUSP5*	2.32135929577311
2	*NFKBIA*	1.90445223426198
3	*EGR2*	1.42742277816002
4	*NR4A1*	1.0766695955654
5	*TGFB1*	1.02420133149033

Supplementary Table 5. Differential immune cell infiltration levels between AR and control groups in the GSE75011 dataset.

ImmuneCell	Mean (AR)	Mean (Control)	P-value
Dendritic cells activated	0.00225	0.00598	0.000247
Macrophages M2	0.00559	0.0169	0.000191
NK cells activated	0.00542	0.00142	0.0138
T cells CD4 memory activated	0.0231	0.0317	0.00625
T cells CD4 naive	0.136	0.0792	0.0203

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]

Supplementary Table 6. Molecular docking parameters.

Docking	Center	Docking size
BOVINECARTILAGE	32,29,4	21,21,21
IDOXURIDINE	62,65,15	27,19,19
TRETINOIN	62,65,15	25,25,25

Supplementary Table 7. Relative mRNA expression levels of *EGR2* and *TGFB1* and corresponding raw *P* values in AR and control groups.

Gene	Normal	AR	Raw p-value
EGR2	1±0.2144	0.6526±0.1674	0.0213
TGFB1	1±0.1518	0.6930±0.1807	0.0196

References:

1. Wen S, Li F, Tang Y, et al. MIR222HG attenuates macrophage M2 polarization and allergic inflammation in allergic rhinitis by targeting the miR146a-5p/TRAF6/NF-kappaB axis. Front Immunol. 2023;14:1168920
2. Okmen HZ, Celiksoy MH, Topal E. The effect of serum vitamin D level on allergic rhinitis symptoms in children. Pediatr Allergy Immunol Pulmonol. 2021;34:132-40
3. Wise SK, Damask C, Roland LT, et al. International consensus statement on allergy and rhinology: allergic rhinitis—2023. Int Forum Allergy Rhinol. 2023;13:293-859
4. Chen X, Wang R, Meng W, Zhang X. Exploration of the molecular mechanism of FUZI (Aconiti Lateralis Radix Praeparata) in allergic rhinitis treatment based on network pharmacology. Med Sci Monit. 2020;26:e920872
5. Fan K, Zhou S, Jin L, et al. Identification of key genes and the pathophysiology associated with allergen-specific immunotherapy for allergic rhinitis. BMC Immunol. 2023;24:19
6. Breiteneder H, Peng Y, Agache I, et al. Biomarkers for diagnosis and prediction of therapy responses in allergic diseases and asthma. Allergy. 2020;75:3039-68.
7. Nur Husna SM, Tan HT, Md Shukri N, Mohd Ashari NS, Wong KK. Allergic rhinitis: a clinical and pathophysiological overview. Front Med (Lausanne). 2022;9:874114
8. Drazdauskaite G, Layhadi JA, Shamji MH. Mechanisms of allergen immunotherapy in allergic rhinitis. Curr Allergy Asthma Rep. 2021;21:2
9. Rollins BJ. Chemokines. Blood. 1997;90:909-28
10. Cambier S, Gouwy M, Proost P. The chemokines CXCL8 and CXCL12: Molecular and functional properties, role in disease and efforts towards pharmacological intervention. Cell Mol Immunol. 2023;20:217-51
11. Chen Y, Liu S, Wu L, et al. Epigenetic regulation of chemokine (CC-motif) ligand 2 in inflammatory diseases. Cell Prolif. 2023;56:e13428
12. Jin J, Lin J, Xu A, et al. CCL2: An important mediator between tumor cells and host cells in tumor microenvironment. Front Oncol. 2021;11:722916
13. Hanna A, Frangogiannis NG. Inflammatory cytokines and chemokines as therapeutic targets in heart failure. Cardiovasc Drugs Ther. 2020;34:849-63
14. Moadab F, Khorramdelazad H, Abbasifard M. Role of CCL2/CCR2 axis in the immunopathogenesis of rheumatoid arthritis: Latest evidence and therapeutic approaches. Life Sci. 2021;269:119034
15. Li Z, Yu S, Jiang Y, Fu Y. Chemokines and chemokine receptors in allergic rhinitis: From mediators to potential therapeutic targets. Eur Arch Otorhinolaryngol. 2022;279:5089-95
16. Mishra T, Sasanka KK, Sudha Ty S, et al. Emerging novel biomarkers in allergic rhinitis: A narrative review. Cureus. 2025;17:e84705
17. Jiang Y, Hu W, Cai Z, et al. Peripheral multiple cytokine profiles identified CD39 as a novel biomarker for diagnosis and reflecting disease severity in allergic rhinitis patients. Mediators Inflamm. 2023;2023:1-11
18. Hu Y, Xiao M, Zhang D, et al. Comprehensive analysis of chemokines family and related regulatory ceRNA network in lung adenocarcinoma. Heliyon. 2022;8:e11399
19. Ritchie ME, Phipson B, Wu D, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015;43:e47
20. Li J, Wang H, Cao F, Cheng Y. A bioinformatics analysis for diagnostic roles of the E2F family in esophageal cancer. J Gastrointest Oncol. 2022;13:2115-31
21. Hanzelmann S, Castelo R, Guinney J. GSVA: Gene set variation analysis for microarray and RNA-seq data. BMC Bioinformatics. 2013;14:7
22. Langfelder P, Horvath S. WGCNA: An R package for weighted correlation network analysis. BMC Bioinformatics. 2008;9:559
23. Kanehisa M, Furumichi M, Sato Y, et al. KEGG: Biological systems database as a model of the real world. Nucleic Acids Res. 2025;53:D672-77
24. Kanehisa M. Toward understanding the origin and evolution of cellular organisms. Protein Sci. 2019;28:1947-51
25. Yu G, Wang L, Han Y, He Q. clusterProfiler: An R package for comparing biological themes among gene clusters. OMICS. 2012;16:284-87
26. Chin C, Chen S, Wu H, et al. cytoHubba: Identifying hub objects and subnetworks from complex interactome. BMC Syst Biol. 2014;8(Suppl. 4):S11
27. Zhang Z, Zhao Y, Canes A, et al. Predictive analytics with gradient boosting in clinical medicine. Ann Transl Med. 2019;7:152
28. Robin X, Turck N, Hainard A, et al. pROC: An open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics. 2011;12:77
29. Alderden J, Pepper GA, Wilson A, et al. Predicting pressure injury in critical care patients: A machine-learning model. Am J Crit Care. 2018;27:461-68
30. Liu T, Li R, Huo C, et al. Identification of CDK2-related immune forecast model and ceRNA in lung adenocarcinoma, a pan-cancer analysis. Front Cell Dev Biol. 2021;9:682002
31. Zhang J, Wang Z, Zhang X, et al. Large-scale single-cell and bulk sequencing analyses reveal the prognostic value and immune aspects of CD147 in pan-cancer. Front Immunol. 2022;13:810471
32. Hadighi R, Heidari A, Fallah P, et al. Key plasma microRNAs variations in patients with Plasmodium vivax malaria in Iran. Heliyon. 2022;8:e09018
33. Bousquet J, Anto JM, Bachert C, et al. Allergic rhinitis. Nat Rev Dis Primers. 2020;6:95
34. Zhu J, Cai M, Lei G, Cao Z. Exploring putative links between gut microbiota and allergic rhinitis: Insights from Mendelian randomization and multi-transcriptome integration. AMB Express. 2025;15:174

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]

35. Li J, Wang J, Lu Y, et al. Single-cell RNA sequencing reveals immune cell alterations in patients with allergic rhinitis treated with Peiyuan Tong-qiao decoction. Biochem Biophys Rep. 2025;44:102325
36. Joerger AC, Fersht AR. The p53 pathway: Origins, inactivation in cancer, and emerging therapeutic approaches. Annu Rev Biochem. 2016;85:375-404
37. Yue J, Lopez JM. Understanding MAPK signaling pathways in apoptosis. Int J Mol Sci. 2020;21:2346
38. Dai R, Wang Z, Heid B, et al. EGR2 deletion suppresses anti-DsDNA autoantibody and IL-17 production in autoimmune-prone B6/lpr mice: A differential immune regulatory role of EGR2 in B6/lpr versus normal B6 mice. Front Immunol. 2022;13:917866
39. Dai R, Heid B, Xu X, et al. EGR2 is elevated and positively regulates inflammatory IFNγ production in lupus CD4+ T cells. BMC Immunol. 2020;21:41
40. Taefehshokr N, Miao T, Symonds ALJ, et al. Egr2 regulation in T cells is mediated through IFNγ/STAT1 and IL-6/STAT3 signalling pathway. Pathol Res Pract. 2020;216:153259
41. Rübsamen D, Blees JS, Schulz K, et al. IRES-dependent translation of egr2 is induced under inflammatory conditions. RNA. 2012;18:1910-20
42. Li W, He S, Tan J, et al. Transcription factor EGR2 alleviates autoimmune uveitis via activation of GDF15 to modulate the retinal microglial phenotype. Proc Natl Acad Sci U S A. 2024;121:e2316161121
43. Myouzen K, Kochi Y, Shimane K, et al. Regulatory polymorphisms in EGR2 are associated with susceptibility to systemic lupus erythematosus. Hum Mol Genet. 2010;19:2313-20
44. Dong F, Tan J, Zheng Y. Chlorogenic acid alleviates allergic inflammatory responses through regulating Th1/Th2 balance in ovalbumin-induced allergic rhinitis mice. Med Sci Monit. 2020;26:e923358
45. Zhu X, Lu M, Chen R, et al. Polymorphism -509C/T in TGFB1 promoter is associated with increased risk and severity of persistent allergic rhinitis in a Chinese population. Am J Rhinol Allergy. 2020;34:597-603
46. Chaker AM, Zissler UM, Poulos N, et al. Activin-A is a pro-inflammatory regulator in type-2-driven upper airway disease. Int Arch Allergy Immunol. 2018;176:15-25
47. Elkoshi Z. TGF-beta, IL-1beta, IL-6 levels and TGF-beta/Smad pathway reactivity regulate the link between allergic diseases, cancer risk, and metabolic dysregulations. Front Immunol. 2024;15:1371753
48. Oh SA, Li MO. TGF-beta: Guardian of T cell function. J Immunol. 2013;191:3973-79
49. Yang H, Kim H, Park J, et al. Apigenin alleviates TGF-beta1-induced nasal mucosa remodeling by inhibiting MAPK/NF-κB signaling pathways in chronic rhinosinusitis. PLoS One. 2018;13:e0201595
50. Haque TT, Frischmeyer-Guerrerio PA. The role of TGFbeta and other cytokines in regulating mast cell functions in allergic inflammation. Int J Mol Sci. 2022;23:10864
51. Long B, Zhou S, Gao Y, et al. Tissue-resident memory T cells in allergy. Clin Rev Allergy Immunol. 2024;66:64-75
52. Wang R, Liang J, Wang Q, et al. m6A mRNA methylation-mediated MAPK signaling modulates the nasal mucosa inflammatory response in allergic rhinitis. Front Immunol. 2024;15:1344995
53. Zhang L, Meng W, Chen X, et al. MiR-150-5p regulates the functions of type 2 innate lymphoid cells via the ICAM-1/p38 MAPK axis in allergic rhinitis. Mol Cell Biochem. 2022;477:1009-22
54. Chandra A, Lan S, Zhu J, et al. Epidermal growth factor receptor (EGFR) signaling promotes proliferation and survival in osteoprogenitors by increasing early growth response 2 (EGR2) expression. J Biol Chem. 2013;288:20488-98
55. Gregory KJ, Morin SM, Schneider SS. Regulation of early growth response 2 expression by secreted frizzled related protein 1. BMC Cancer. 2017;17:473
56. Wu W, Li J, Chen S, Ouyang S. The airway neuro-immune axis as a therapeutic target in allergic airway diseases. Respir Res. 2024;25:83
57. Hao Y, Wang B, Zhao J, et al. Identification of gene biomarkers with expression profiles in patients with allergic rhinitis. Allergy Asthma Clin Immunol. 2022;18:20
58. Zhang J, Xu M, Zhou W, et al. Deficiency in the anti-apoptotic protein DJ-1 promotes intestinal epithelial cell apoptosis and aggravates inflammatory bowel disease via p53. J Biol Chem. 2020;295:4237-51
59. Wang J, Chang CY, Yang X, et al. p53 suppresses MHC class II presentation by intestinal epithelium to protect against radiation-induced gastrointestinal syndrome. Nat Commun. 2024;15:137
60. Carrà G, Lingua MF, Maffeo B, et al. P53 vs NF-κB: The role of nuclear factor-kappa B in the regulation of p53 activity and vice versa. Cell Mol Life Sci. 2020;77:4449-58
61. Yokota I, Sasaki Y, Kashima L, et al. Identification and characterization of early growth response 2, a zinc-finger transcription factor, as a p53-regulated proapoptotic gene. Int J Oncol. 2010;37:1407-16
62. Overstreet JM, Samarakoon R, Meldrum KK, Higgins PJ. Redox control of p53 in the transcriptional regulation of TGF-β1 target genes through SMAD cooperativity. Cell Signal. 2014;26:1427-36
63. Williams CMM, Rahman S, Hubeau C, Ma HL. Cytokine pathways in allergic disease. Toxicol Pathol. 2012;40:205-15
64. Choi A, Jung YW, Choi H. The extrinsic factors important to the homeostasis of allergen-specific memory CD4 T cells. Front Immunol. 2022;13:1080855
65. Gong D, Shi W, Yi S, et al. TGFβ signaling plays a critical role in promoting alternative macrophage activation. BMC Immunol. 2012;13:31
66. Miao T, Symonds ALJ, Singh R, et al. Egr2 and 3 control adaptive immune responses by temporally uncoupling expansion from T cell differentiation. J Exp Med. 2017;214:1787-808
67. Szymanski L, Skopek R, Palusinska M, et al. Retinoic acid and its derivatives in skin. Cells. 2020;9:2600
68. Sousa-Pinto B, Vieira RJ, Brozek J, et al. Intranasal antihistamines and corticosteroids in allergic rhinitis: A systematic review and meta-analysis. J Allergy Clin Immunol. 2024;154:340-54

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]