Citation: Clin Transl Sci (2019) XX, 1–10; doi:10.1111/cts.12698

ARTICLE

Identification of Susceptibility Genes to Allergic Rhinitis by Gene Expression Data Sets

Kai Xue[1], Jingpu Yang[1], Yin Zhao[1], Jinzhang Cheng[1] and Zonggui Wang[1,*]

As an extremely prevalent disease worldwide, allergic rhinitis (AR) is a condition characterized by chronic inflammation of the nasal mucosa. To identify the finer molecular mechanisms associated with the AR susceptibility genes, differentially expressed genes (DEGs) in AR were investigated. The DEG expression and clinical data of the GSE19187 data set were used for weighted gene co-expression network analysis (WGCNA). After the modules related to AR had been screened, the genes in the module were extracted for Gene Ontology and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis, whereby the genes enriched in the KEGG pathway were regarded as the pathway-genes. The DEGs in patients with AR were subsequently screened out from GSE19187, and the sensitive genes were identified in GSE18574 in connection with the allergen challenge. Two kinds of genes were compared with the pathway-genes in order to screen the AR susceptibility genes. Receiver operating characteristic (ROC) curve was plotted to evaluate the capability of the susceptibility genes to distinguish the AR state. Based on the WGCNA in the GSE19187 data set, 10 co-expression network modules were identified. The correlation analyses revealed that the yellow module was positively correlated with the disease state of AR. A total of 89 genes were found to be involved in the enrichment of the yellow module pathway. Four genes (*CST1*, *SH2D1B*, *DPP4*, and *SLC5A5*) were upregulated in AR and sensitive to allergen challenge, whose potentials were further confirmed by ROC curve. Taken together, *CST1*, *SH2D1B*, *DPP4*, and *SLC5A5* are susceptibility genes to AR.

Study Highlights

WHAT IS THE CURRENT KNOWLEDGE ON THE TOPIC?

☑ Allergic rhinitis (AR) is one of the most prevalent chronic conditions around the world with its pathogenesis arising due to both genetic and environmental factors. The increased risk of AR is associated with genes that regulate immune responses, such as *TLR4* and *CD14*.

WHAT QUESTION DID THIS STUDY ADDRESS?

☑ What are the susceptibility genes to AR?

WHAT DOES THIS STUDY ADD TO OUR KNOWLEDGE?

☑ Four genes (*CST1*, *SH2D1B*, *DPP4*, and *SLC5A5*) were upregulated in AR and sensitive to allergen challenge, suggesting that *CST1*, *SH2D1B*, *DPP4*, and *SLC5A5* are susceptibility genes to AR.

HOW MIGHT THIS CHANGE CLINICAL PHARMACOLOGY OR TRANSLATIONAL SCIENCE?

☑ Investigation of susceptibility genes in AR and their functions yields a better understanding of mechanisms underlying AR and may have potentially important therapeutic implications in the treatment of AR.

Allergic rhinitis (AR) is one of the most prevalent chronic conditions around the world, with people of all ages exhibiting various symptoms, such as repetitive sneezing, nasal itching, rhinorrhea, as well as nasal obstruction.[1] AR has been linked with diminished quality of life, reduced sleep quality, and cognitive function, as well as heightened irritability and fatigue.[2] Although valiant efforts have been made to alleviate the symptoms of AR as well as to identify the finer molecular mechanisms associated with its pathological changes of the nasal mucosa, the treatment of AR remains a challenging task, highlighting the importance of identifying the key molecular and genetic entities that trigger AR pathologies, which may ultimately lead us to discover new targets for the treatment of AR.[3] More recently, genes that regulate immune responses have been shown to contribute to the increased risk of AR, with the relationship among AR, *TLR4*, and *CD14* implicated in the occurrence of AR.[4] However, the current evidence cannot fully explain the high incidence of AR.

At present, gene therapy has been identified as a potential method for the treatment of allergic airway diseases, including seasonal AR, so it is critical to retrieve candidate target molecules for the treatment of AR.[5] The Gene Expression Omnibus (GEO) database provides a flexible and open design for submitting, storing, and retrieving heterogeneous data sets from high-throughput gene

[1]Department of Otolaryngology-Head and Neck Surgery, The Second Hospital of Jilin University, Changchun, China. *Correspondence: Zonggui Wang (wangzonggui1965@126.com)
Received: March 29, 2019; accepted: August 6, 2019. doi:10.1111/cts.12698

expression and genomic hybridization experiments.[6] The Kyoto Encyclopedia of Genes and Genomes (KEGG; http://www.genome.jp/kegg/), a database of biological systems, integrates genomic, chemical, and systemic functional information.[7] Tremendous efforts have been made to identify differentially expressed genes (DEGs) in AR using microarray data. For instance, the Gene Ontology (GO) Biological Process (BP) and KEGG pathway enrichment analyses were adopted in a previous study, which found that FOS, JUN, and CEBPD may exert vital functions during the progression of seasonal AR.[8] Meanwhile, another study also observed that CST1, CLC, and STAT1 were associated with AR by utilizing the GEO database and KEGG analysis.[9]

The gene expression data sets (GSE19187 and GSE18574) uploaded from the GEO database, were used in the current study to perform a series of microarray analyses to identify novel AR targets by detected the biological function of DEGs involved in progression of AR. A weighted gene co-expression network analysis (WGCNA) was used to identify the gene modules associated with AR followed by identification of the DEGs between patients with AR and healthy individuals. Moreover, GO and ClueGO pathway enrichment analysis was performed for the genes in the yellow module that was positively correlated with AR. The intersection among DEGs, sensitive genes to allergen challenge, and genes in the yellow module was verified as susceptibility genes to AR. Additionally, potentials of susceptibility genes for prediction of the disease state were also investigated based on the receiver operating characteristic (ROC) curve. The purpose of this study was to identify the potential target genes and explore the underlying mechanisms in the pathogenesis of AR.

MATERIALS AND METHODS

Microarrays

Gene expression profile and gene annotation files of AR-related microarrays GSE19187 and GSE18574 were downloaded from the GEO database (https://www.ncbi.nlm.nih.gov/geo/).[10] The GSE19187 data set consisted of 38 samples, was classified into four categories, namely: 14 isolated rhinitis (R), six rhinitis with uncontrolled asthma, seven rhinitis with controlled asthma, and 11 healthy subjects (control). Only healthy individuals and patients with AR were used for the subsequent WGCNA and DEG screening process. The sequencing tissue in the microarray was epithelial cell, and all the patients were allergic to dust mite.[11] GSE19187 data referred to data of healthy individuals and patients with AR. The gene annotation platform of GSE19187 database was GPL6244 (HuGene-1_0-st) Affymetrix Human Gene 1.0 ST Array (transcript (gene) version). GSE18574 was comprised of samples from three patients with AR; each separated into allergen-challenged samples and unchallenged controls, among which grass pollen extract was the allergen of allergen-challenged for screening the allergen-challenged susceptibility genes in patients with AR. CD4+ cells were used as sequencing cells.[12] The gene annotation platform of GSE18574 was GPL2507 Sentrix Human-6 Expression BeadChip.

WGCNA

WGCNA was applied to calculate the correlation efficient, build hierarchical clustering tree, and divide the genes with high co-expression into the same module based on their respective expression levels. Based on the data of the healthy individuals and patients with AR from the GSE19187 data set, a co-expression network module was constructed with R software and the WGCNA package utilized.[13] In order to satisfy the premise of scale-free network distribution among genes in the co-expression network, an adjacency matrix weight parameter β value was set at 1–20 during the current study. The One-step network construction and module detection method were used to construct co-expression network module.

The correlation analysis and identification of AR-related modules

After the genes had been divided into different modules by the WGCNA analysis, the correlation among different modules was calculated with the correlation heatmap used to help identify visualize the correlation between all the modules. The GSE19187 data set contains the clinical information, including the disease state, sex, and age of the samples. In an attempt to further investigate the correlation between the module and the disease, the correlation between the module and the clinical factors was evaluated to detect the module with the greatest susceptibility to AR, with the WGCNA package in R software (Vienna, Austria) used for analysis.

Enrichment analysis

In order to assess their biological function, the module genes were extracted and the function enrichment analysis of GO BP, cellular component, and molecular function was carried out via clusterProfiler package[14] in R software after extraction of the AR-related modules utilized. The Benjamini-Hochberg multiple hypothesis test was used to correct the enrichment results with adjusted $P < 0.05$ set as the threshold. ClueGO, a CytoScape plug-in that allows functional annotation of genes, was used for signaling pathway enrichment analysis of key module genes to further visualize the interrelation networks between genes and pathways. The gene enriched into the pathway was defined as the module pathway-gene.

DEG screening

The DEGs of AR were screened out using R software, based on the microarray data of the healthy individuals as well as the patients with AR in GSE19187. The DEGs that were easily affected during the allergen challenged were identified from the GSE18574 microarray accordingly. The affy package in R software[15] was used to pre-process the expression of the microarray. The Limma package[16] was applied for DEG screening, and the genes with an adjusted P value < 0.05 and $|\log2 \text{ fold change}| > 1.0$ were selected as the DEGs. The P value was corrected using the false discovery rate method. The heatmap of DEGs was plotted with the pheatmap package (https://cran.r-project.org/web/packages/pheatmap/index.html). By comparing the DEGs with the module pathway-gene, the intersection genes were identified, which were the module pathway-genes that were differentially expressed in AR and easily affected by allergen challenge were the key genes in AR.

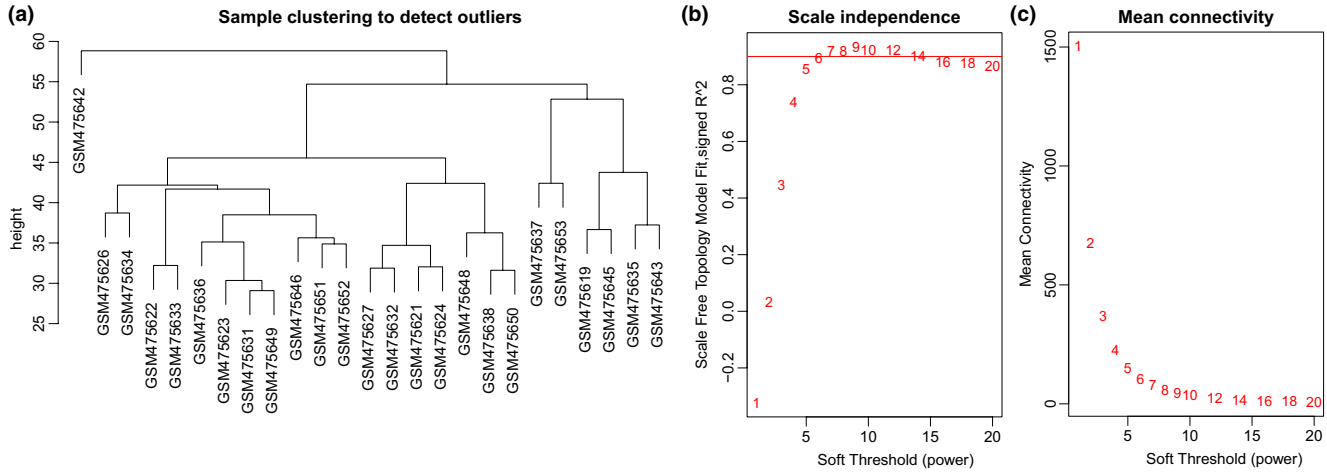

Figure 1 WGCNA processing for AR data set GSE19187. (**a**) Sample clustering dendrogram from WGCNA; (**b, c**) power value for the adjacency matrix in WGCNA, where the red line signals 0.9 on the vertical axis. AR, Allergic rhinitis; WGCNA, weighted gene co-expression network analysis.

ROC curve analysis

In order to investigate the effect associated with gene expression on the disease state of AR, the ROC curve was plotted based on the gene expression data from the GSE19187 and the state of the sample, with the gene expression accuracy assessed accordingly. The pROC package[17] in R software was used to plot ROC curves. The weight of each gene was determined using the entropy weight method, with the ROC curve of four genes subsequently plotted. The area under the curve (AUC) obtained from an ROC curve analysis was utilized to assess the biomarker accuracy. The sensitivity and specificity for each gene was separately calculated in addition to the cutoff values acquired from the ROC curve.

Statistical analysis

The statistical package R version 3.5.1 was applied to analyses gene expression data. The "affy" and "limma" were used to analyze the standardized correction differences of microarray data. The "WGCNA" package was utilized for gene co-expression module analysis, and the "clusterprofiler" package was used for functional enrichment analysis of the DEGs. The "pROC" package was used to plot an ROC curve and calculate the AUC value. An adjusted $P < 0.05$ value following false discovery rate correction was considered to be statistically significant.

RESULTS

WGCNA analysis

The flow diagram of our protocol is shown in **Figure S1**. GSE19187 expression data and annotation files were downloaded from the GEO database. In order to construct the co-expression module, the expression data of both the healthy individuals as well as the patients with AR were extracted from GSE19187, after which the expression profiles were annotated to 19,976 genes. The SD value of each gene was calculated, followed by ranking of the top 25% of genes (4,994 genes) in descending order followed by WGCNA analysis. In order to evaluate the outliers of the samples, sample clustering methods were used based on the distribution of the expression values of the samples (**Figure 1a**), with no significant difference detected in the samples included in the WGCNA. A scale-free topology index and mean connectivity were applied during the current study to determine the soft-threshold in WGCNA. The higher the scale-free topology index value was indicative of a strong probability of a nonscale feature. Power = 5 was selected in the event of the correlation coefficient between log (*k*) and log P (*k*) reached 0.9 for the first time (**Figure 1b**).

Module selection

After determination of the optimal soft-threshold = 5, both the relatively balanced scale independence and the mean connectivity of WGCNA were ensured accordingly. Within the gene co-expression network module recognition, the maximum number of genes processed by computer was maxBlockSize = 6,000, the minimum number of genes of each module was minModuleSize = 30, and the module merge threshold was set as mergeCutHeight = 0.25. The 10 identified modules are depicted in **Figure 2a**, whereby the different modules were marked using different colors. The modules in gray were reflective of genes that were not divided into any other module, and a smaller number was considered to be representative of a superior result. The turquoise, blue, and brown were found to have the largest number of genes, with 1,660, 1,274, and 480 genes, respectively. The detailed distribution of genes in each module can be seen illustrated in **Figure 2b**.

The correlation analysis between modules and the recognition of the key module

The co-expression of the genes of the modules in blue, turquoise, brown, and red was relatively high. The correlation between the module genes is illustrated in **Figure 3a**. In order to further explore the relationship between modules, the correlation between modules was calculated accordingly (**Figure 3b**). A hierarchical clustering tree (**Figure 3c**) was plotted to observe the similarity between the modules, after which 10 co-expression modules from WGCNA were

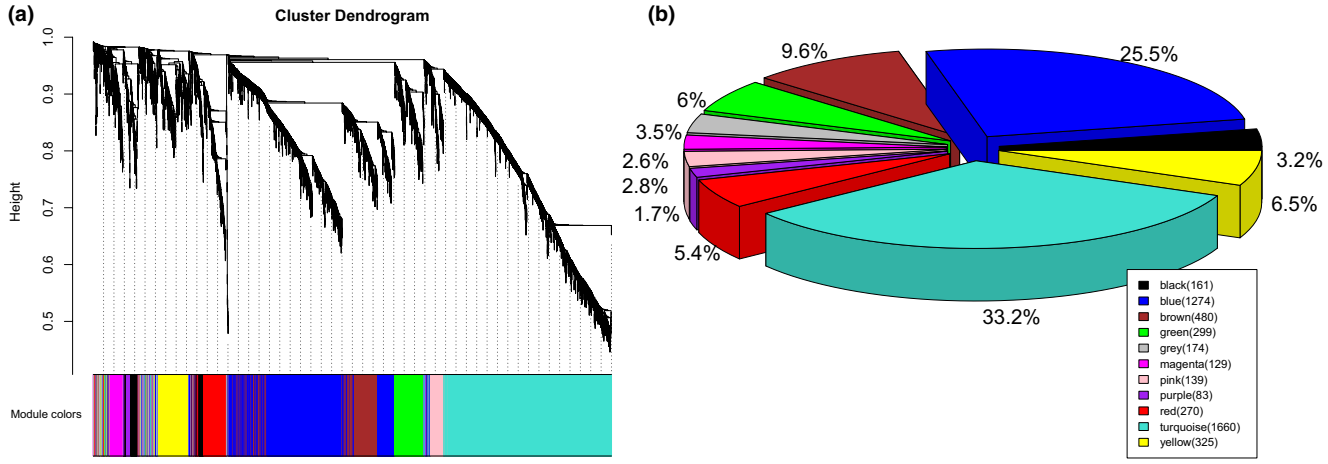

Figure 2 The number of modules and genes from WGCNA. (**a**) The cluster dendrogram of genes in GSE19187 with each branch representing a gene, and each color representative of a co-expression module; (**b**) the number of co-expression module genes identified by WGCNA, and the number in parentheses of the legend represents the number of genes in each module. WGCNA, weighted gene co-expression network analysis.

divided into 2 categories, these modules clustered into the same large class had a certain similarities in terms of their gene expression trends, which was also detected based on our analysis of the correlation heatmap in **Figure 3b**. In order to obtain the gene modules closely related to AR, the relevant clinical information of the sample was extracted from the microarray, and the correlation between the above 10 different color modules and the 3 clinical characteristics of age, sex, and disease state were analyzed. The heatmap of the module and the clinical correlation is demonstrated in **Figure 3d**. We observed that the yellow module shared the closest correlation to AR, which was identified to be a positive association. The correlation coefficient reached 0.54, which was considered to be an indication that the genes in this module were highly likely to be positively affected by the development of the disease. The gray module was found to be negatively correlated with AR, and the genes in the gray module could not be divided into any of the modules. Thus, based on the results obtained, we concluded that the genes in this module may not directly influence the development of the disease. In addition, the correlation coefficient between the magenta module and the disease status was found to be −0.3, suggesting that the genes in this module could be negatively correlated with disease status, which was still notably lower than the yellow module. The higher the correlation with the disease, the greater the potential for disease, hence, highlighting the yellow module as a key AR module. The core genes closely related to AR were obtained through the application of an in-depth analysis of the genes in the module.

GO enrichment analysis

Based on the aforementioned correlation analysis of the WGCNA co-expression modules and clinical features, the yellow module has the highest correlation with AR. In order to identify the biological functions of the genes involved in the module, 325 genes were extracted from the yellow module. The GO enrichment analysis was performed using R software. The genes in the yellow module were then markedly enriched in BP terms, including extracellular matrix organization, cornification, extracellular structure organization, epidermis development, regulation of peptidase activity (**Figure 4a**), and also enriched in cellular component terms, such as extracellular matrix, proteinaceous extracellular matrix, extracellular matrix component, basement membrane, and the basolateral plasma membrane (**Figure 4b**), all of which were associated with the cell matrix. Regarding the molecular function terms, genes in the yellow module were predominantly enriched in peptidase regulator activity, endopeptidase inhibitor activity, endopeptidase regulator activity, peptidase inhibitor activity, and growth factor binding (**Figure 4c**). The obtained results indicated that the aforementioned items were strongly involved in the development of AR. The number of items associated with peptidase regulation occurred frequently in GO analysis, suggesting a higher association with AR. Reported studies have shown that the regulation of peptidase activity was closely related to AR, which indicated that our previous analysis was reasonable and credible.

ClueGO pathway enrichment analysis

With an aim to understand the information related to the signaling pathways involved in the development of AR, enrichment analysis of KEGG signaling pathway was conducted on the gene in the co-expression yellow module using the CytoScape plug-in ClueGO. Following that, the correlation between the gene and the pathway was analyzed in an attempt to elucidate the relevant details of the metabolic pathway in AR. As depicted in **Figure 5**, the genes in the yellow module were markedly enriched in pathways such as arachidonic acid (AA) metabolism, extracellular matrix (ECM)-receptor interaction, p53 signaling pathway, focal adhesion, mucin type O-glycan biosynthesis, and amoebiasis. There were 89 genes identified to be involved in the enriched pathway in the yellow module, and our observations revealed that the yellow module was positively related to the disease state. The results suggested that the aforementioned genes were

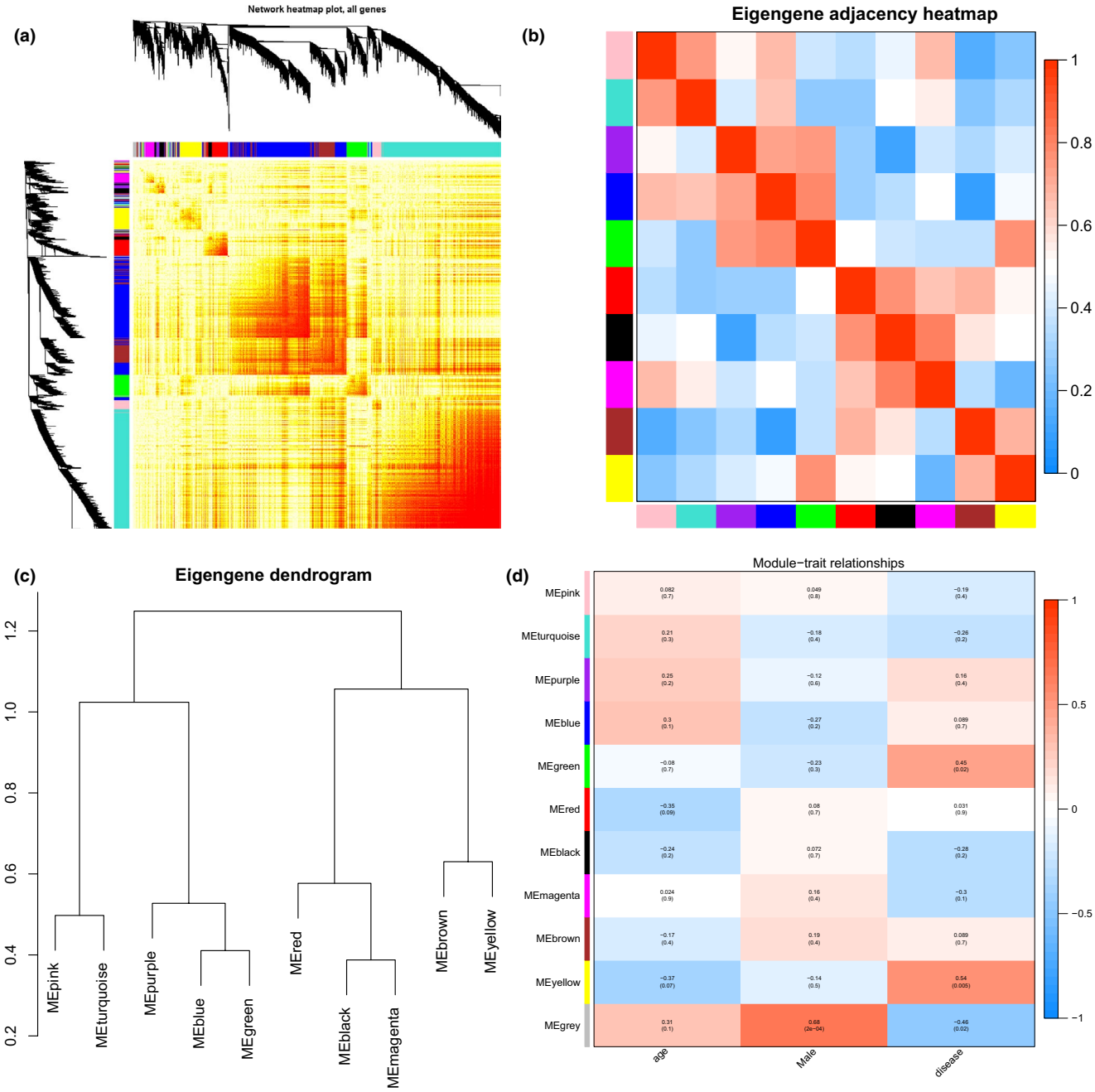

Figure 3 The correlation between WGCNA modules. (**a**) The heatmap of co-expression gene module, where deeper color is reflective of a stronger correlation; (**b**) the heatmap of co-expression modules, where red for positive correlation and blue for negative correlation; (**c**) the clustering dendrogram for WGCNA co-expression modules in which shorter distance indicates higher similarity; (**d**) the correlation between WGCNA modules and the clinical features of the sample. Each row corresponds to a module, and the column corresponds to different clinical features. The numbers in the figure represent the correlation between the corresponding module and the clinical feature with the *P* value displayed in parentheses. The red represents a positive correlation, whereas blue indicates a negative correlation). WGCNA, weighted gene co-expression network analysis.

involved in the regulation of the AP through the pathway, which were subsequently used for the module pathway-gene analyses.

Identification of DEGs

The occurrence of AR has been strongly correlated with a large variety of allergens. To further our understanding regarding the genes that are susceptible to allergens and had significant differences in AR, R software was used to analyze and screen DEGs in AR with the data from healthy individuals and patients with AR obtained from GSE19187 dataset, and 24 DEGs were screened out. Only two genes were found to be poorly expressed in AR when compared with the findings among the healthy individuals, whereas

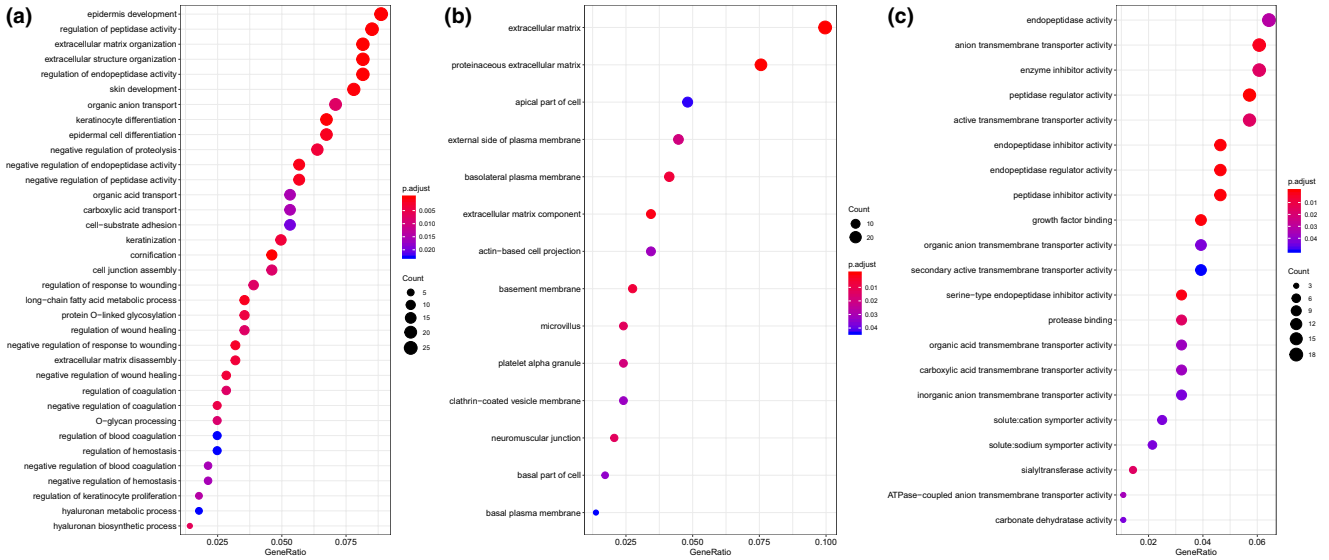

Figure 4 GO enrichment analysis of genes in the yellow module. (**a**) Significant BP terms; (**b**) significant CC terms; (**c**) significant MF terms. Vertical axis represents GO terms. Circle size represents the number of genes significantly enriched in each GO term. *P* values are marked in different colors. BP, biological process; CC, cellular component; GO, gene ontology; MF, molecular function.

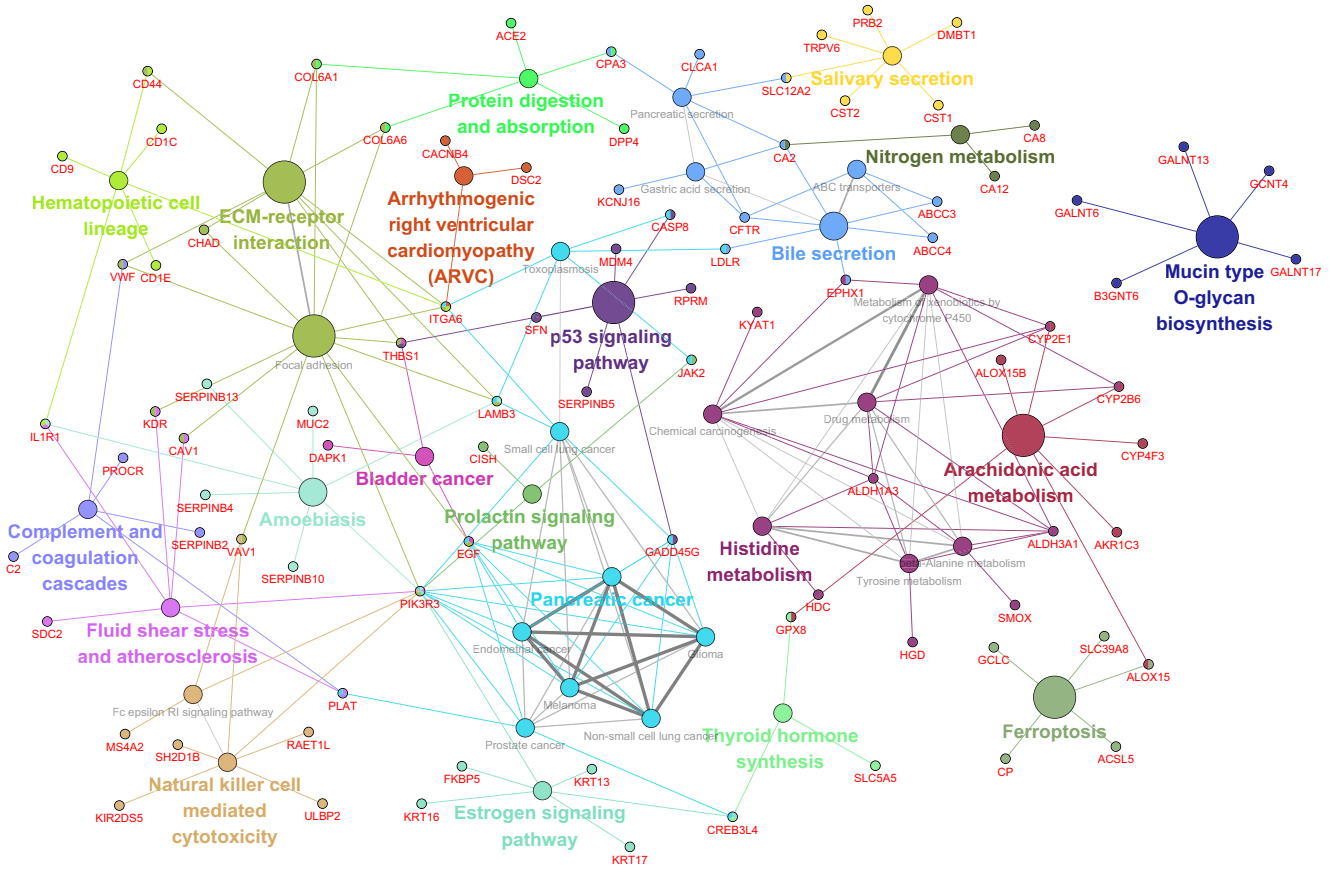

Figure 5 ClueGO pathway enrichment analysis of genes in the yellow module. The smallest circle labeled with the red word represents the enriched module gene, and the other larger circles represent the ClueGO signaling pathway that was enriched by the module gene. GO, gene ontology.

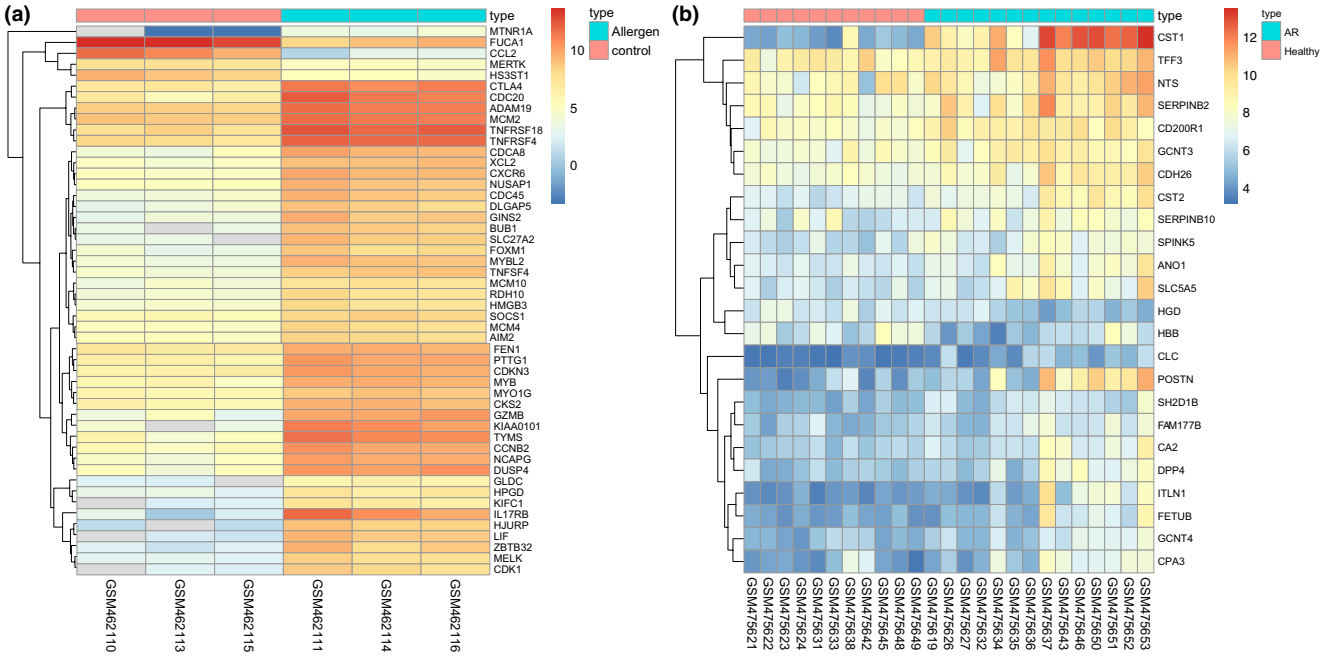

Figure 6 Heatmap of DEGs between patients with AR and the healthy individuals. (**a**) The heatmap of DEGs in GSE19187; (**b**) the heatmap of the top 50 DEGs between allergen challenged and unchallenged patients with AR in GSE18574. Each row represents a sample number; each column represents a single gene. The gradual color change from red to blue represents the changing process from upregulation to downregulation. Allergen, sample challenged by allergen; AR, allergic rhinitis; control, samples free of allergen challenge; DEGs, differentially expressed genes.

the other 22 genes all exhibited high expression levels. The heatmap of DEGs is displayed in **Figure 6a**. Of the 24 genes analyzed, our results suggested a strong likelihood that they are involved in the development of AR. GSE18574 demonstrates data of allergen-challenged and unchallenged patients with AR. Genes that were sensitive to allergen challenge were screened out. In the patients with AR, the expression of 2,218 genes was changed after allergen challenge, therefore, these genes were deemed to be susceptible to the allergen challenge. The heatmap of the top 50 DEGs is illustrated in **Figure 6b**.

Allergen sensitive gene in AR

The three major types of genes were selected through the aforementioned analyses were as follows: 89 pathway-genes in the yellow module were screened through WGCNA; 24 DEGs in the GSE19187 data set expressing differentially in the patients with AR and healthy individuals, and 2,218 sensitive genes to allergen challenge in the GSE18574 data set. The comparison among these three types of genes, revealed there to be five intersecting genes: cystatin SN (CST1), SH2D1B, dipeptidyl peptidase 4 (DPP4), SERPINB2, and SLC5A5 (**Figure 7a**), with the expression of these five genes in AR was higher than that in healthy individuals in the GSE19187 data set (**Figure 7b**). In addition, in the GSE18574 data set, *CST1*, *SH2D1B*, *DPP4*, and *SLC5A5* were also highly expressed after the allergen challenged in patients with AR, with only low expression of SERPINB2 detected (**Figure 7c**). *CST1*, *SH2D1B*, *DPP4*, and *SLC5A5* were expressed differently in AR and easily affected by allergens, highlighting the notable role of these genes in the development of AR. Based on the *CST1*, *SH2D1B*, *DPP4*, and *SLC5A5* expression data in GSE19187 and the disease state of the samples, an ROC curve was plotted to evaluate the significance of these four genes in AR. As depicted in **Figure 7d**, the AUC of these four genes was >0.8, indicating these genes could be used to identify AR. Among these four genes, the AUC value of the *CST1* gene reached 0.985, indicating that this gene was more suitable as a marker gene for AR than the other three genes. With four genes as a gene data set, the weight of each gene was calculated using the entropy weight method. The ROC curve of four gene data sets was constructed, after which the AUC value was calculated accordingly (**Figure 7e**). The results obtained revealed that the AUC value was 0.981 when using the four genes as a gene data set for the diagnosis of AR, highlighting the ability of these four genes as good markers for AR.

DISCUSSION

AR is a complex disorder with its pathogenesis arising due to both genetic and environmental factors results in its pathogenesis.[18] There were ~ 400 million people who have AR in the Asia-Pacific area, which had a tremendous influence on quality of life, performance at school, and in the workplace and caused a huge socio-economic burden, thus, more understanding on AR diagnosis, epidemiology, risk factors, and prevention is urgently needed.[19] At present, the factors that negatively impact nasal conditioning in patients with AR have not been systematically evaluated,

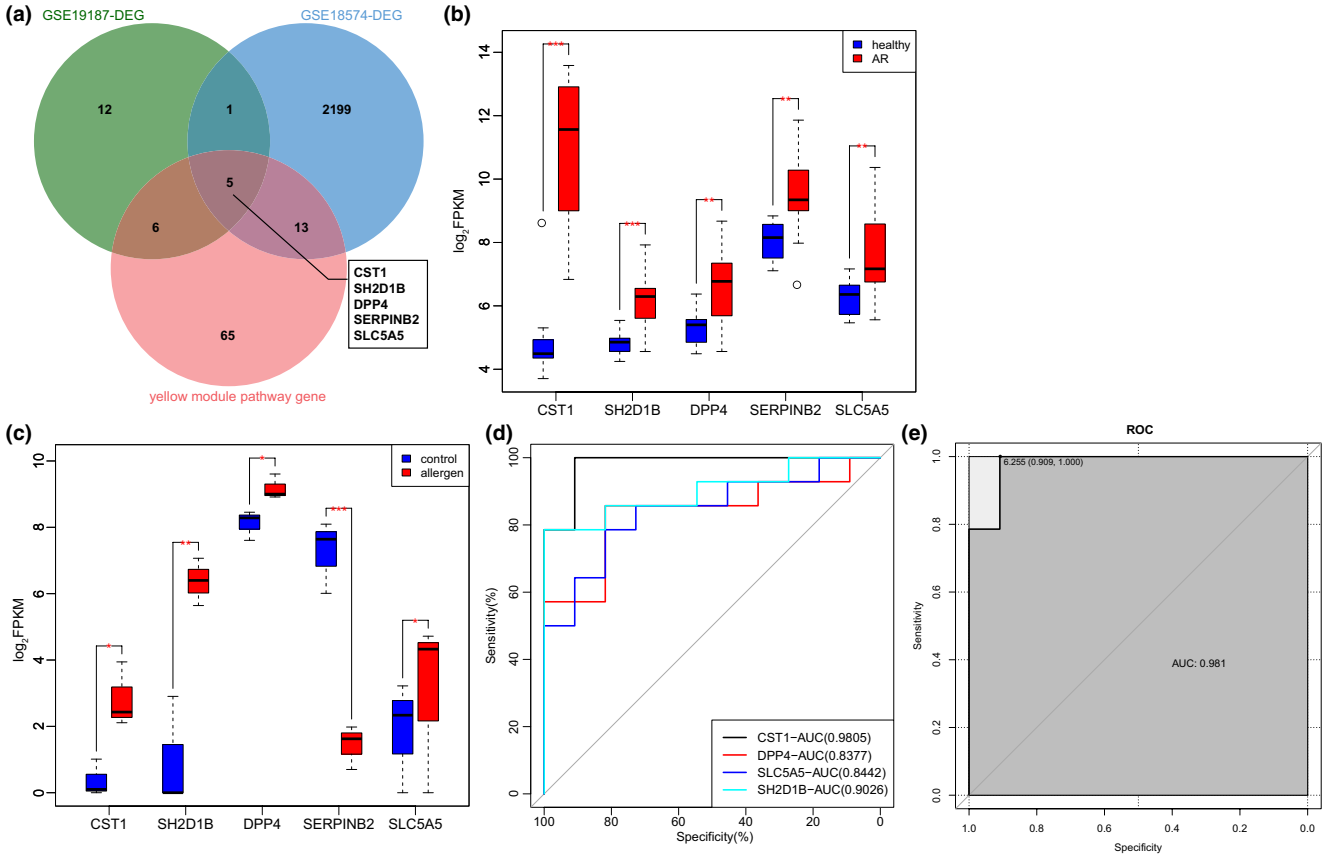

Figure 7 Susceptibility genes to AR were identified. (**a**) Five genes in the intersection of the pathway-genes in yellow module, DEGs in GSE19187 and sensitive gene to allergen challenge in GSE18574; (**b**) the expression of *CST1*, *SH2D1B*, *DPP4*, *SERPINB2*, and *SLC5A5* in GSE19187; (**c**) the expression of *CST1*, *SH2D1B*, *DPP4*, *SERPINB2*, and *SLC5A5* in GSE18574; (**d**) expression of *CST1*, *SH2D1B*, *DPP4*, *SLC5A5*, and ROC curve of AR; (**e**) ROC curve of four genes as gene collection for patients with AR. $^{*}P < 0.05$; $^{**}P < 0.01$; $^{***}P < 0.0001$. AR, allergic rhinitis; DEGs, differentially expressed genes; ROC, receiver operating characteristic.

with accumulating studies highlighting the roles of microRNAs[20] as well as long non-coding RNAs[3] in the disease pathology from a gene expression regulation perspective. In order to elucidate the molecular mechanisms underlying the progression of AR, the current study was set out through the use of WGCNA to identify gene modules associated with progression of AR. A key observation of the current study revealed the DEGs that were particularly sensitive to the allergen challenge in AR (*CST1*, *SH2D1B*, *DPP4*, and *SLC5A5*) were significantly enriched with several GO terms and ClueGO pathways, such as AA metabolism, ECM-receptor interaction, p53 signaling pathway, focal adhesion, mucin type O-glycan biosynthesis, and amoebiasis pathways.

As a member of the type 2 cystatin (CST) superfamily, CST1 has been reported to be capable of inhibiting the proteolytic activities of cysteine proteases and is implicated in progression of several human cancers.[21] CST1 is an endogenous cysteine protease inhibitor that is elevated in nasal epithelia in patients with AR.[22] During natural allergen exposure, CST1 has been reported to be overexpressed in patients with seasonal AR-cryptomeria japonica, and may inhibit cryptomeria japonica allergen-induced histamine release *in vitro*.[23] Furthermore, a previous study ranked CST1 as the number 1 differentially expressed mRNA in AR during an investigation aiming to construct an AR-specific transcriptional regulatory network.[9] The involvement of SH2D1B, also known as Ewing's sarcoma transcript-2, has been implicated in the modulation of signaling the lymphocytic activation molecule family receptor function.[24] SH2D1B, is predominantly expressed in the innate immune cells, such as the macrophages and dendritic cells, whose upregulation has been speculated to help modulate the kinetics of a series of vital proinflammatory cytokine and chemokine responses.[25] DPP4, also referred to CD26, is a type II transmembrane protein that exhibits enhanced levels in an extensive range of metabolic diseases involving diabetes, cardiovascular diseases, and obesity, as well as nonalcoholic fatty liver diseases.[26] Notably, the silencing DPP4 in hepatocytes has been reported to repress the inflammation seen in visceral adipose tissues as well as diminishing the insulin resistance associated with obesity.[27] Evidence has demonstrated the involvement of DPP4 in certain immunological processes, which play a significant role in the lung's allergic responses.[28] The upregulation of DPP4 has been implicated in atopic dermatitis, as well as studies highlighting its involvement in other inflammatory skin diseases as well.[28] More recently, accumulating literature has emerged

providing evidence emphasizing the crucial role of DPP4 in the immune system, especially in the release of T cells.[29] The solute carrier family 5 member 5 (SLC5A5, National Inpatient Sample) gene, located on chromosome 19 (19p13.11), encodes a highly specialized 80–90 kDa transmembrane glycoprotein that helps to regulate the active transport of iodide from the blood into the follicular cells.[30] Guerrieri *et al*.[31] concluded that SLC5A5 is a direct target gene of the p53 signaling pathway and suggested that the DNA-damaging agents-mediated SLC5A5 could hypothetically potentiate its upregulation *in vivo*. Solute-linked carrier has been linked to disease resistance and susceptibility in animals, with research proposing its potential as an essential mechanism in the mediation against intracellular infection.[32] Those findings were largely inconsistent with the findings of the current study, whereby these five genes were all markedly elevated in AR, and displayed sustained high levels following allergen challenge.

AA metabolism has been shown to trigger a large release of a number of inflammatory regulators, including leukotrienes and prostaglandins, which may be involved in numerous inflammatory-associated diseases, such as asthma, rheumatoid arthritis, as well as various types of cancers.[33] Similarly, reports have indicated that AA metabolism and the p53 signaling pathway as the most strongly associated pathways involved with AR. Apart from AA metabolism and the p53 signaling pathway, the DEGs in AR also highlight the role of ECM-receptor interaction. Pathologically, ECM remodeling and infiltration of inflammatory cells are closely correlated to eosinophilic asthma,[34] highlighting its potential in the treatment of AR, a condition related to asthma. More recently, literature has emerged providing findings about focal adhesion regulating cell migration in AR.[35] Interestingly, the DEGs in enchondromas were primarily associated with ECM-receptor interaction, focal adhesion, and amoebiasis as well.[36] Besides, mucin type O-glycan biosynthesis pathway has been previously reported to be significantly enriched in upregulated miRNAs in chronic rhinosinusitis with nasal polyps use, as per the KEGG database.[37] To verify the results of bioinformatics analysis, we used ROC curve to predict the prognostic value of these susceptibility genes, which were significantly correlated with the state of AR.

CONCLUSION

Taken together, the key findings revealed a total of four genes, including *CST1*, *SH2D1B*, *DPP4*, and *SLC5A5* as potential AR biomarkers. The current study provides evidence for the diagnosis and treatment of AR in the future. However, our study has certain limitations, such as the application of only the ROC curve to predict the prognostic value of the susceptibility genes, as well as a small sample size, a lack of validation cohort, and a lack of AR severity data, all of which might influence our results. In addition, the data set is not sufficiently robust, and the reproducibility has not been determined. The prognostic utility of these markers has to be explored in patients with AR and further studies need to be conducted to examine the underlying mechanisms and related pathways of these genes.

Supporting Information. Supplementary information accompanies this paper on the *Clinical and Translational Science* website (www.cts-journal.com).

Figure S1. The flow diagram of our protocol.

Acknowledgments. We would like to acknowledge the reviewers for their helpful comments on this paper.

Funding. No funding was received for this work.

Conflict of interest. The authors declared no competing interests for this work.

Author contributions. J.P.Y., J.Z.C., and K.X. wrote the manuscript. K.X., J.P.Y., Y.Z., and Z.G.W. designed the research. K.X., Y.Z., J.Z.C., and Z.G.W. performed the research. K.X., J.P.Y., and Z.G.W. analyzed the data.

1. Blaiss, M.S. Allergic rhinitis: direct and indirect costs. *Allergy Asthma Proc.* **31**, 375–380 (2010).
2. Brozek, J.L. *et al.* Allergic Rhinitis and its Impact on Asthma (ARIA) guidelines-2016 revision. *J. Allergy Clin. Immunol.* **140**, 950–958 (2017).
3. Ma, Z. *et al.* Identification and functional profiling of differentially expressed long non-coding RNAs in nasal mucosa with allergic rhinitis. *Tohoku J. Exp. Med.* **242**, 143–150 (2017).
4. Seo, J.H. *et al.* Interactions between innate immunity genes and early-life risk factors in allergic rhinitis. *Allergy Asthma Immunol. Res.* **7**, 241–248 (2015).
5. Liu, Y., Shi, J. & Chen, X. Identification of novel targets for seasonal allergic rhinitis during and outside the pollen season by microarray analysis. *Acta Otolaryngol.* **135**, 1330–1336 (2015).
6. Edgar, R., Domrachev, M. & Lash, A.E. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* **30**, 207–210 (2002).
7. Kanehisa, M. *et al.* KEGG for linking genomes to life and the environment. *Nucleic Acids Res.* **36**, D480–D484 (2008).
8. Shi, J. *et al.* Identification of potential crucial gene network related to seasonal allergic rhinitis using microarray data. *Eur. Arch. Otorhinolaryngol.* **274**, 231–237 (2017).
9. Lei, Y., Guo, P., An, J., Guo, C., Lu, F. & Liu, M. Identification of pathogenic genes and upstream regulators in allergic rhinitis. *Int. J. Pediatr. Otorhinolaryngol.* **115**, 97–103 (2018).
10. Barrett, T. *et al.* NCBI GEO: archive for functional genomics data sets – update. *Nucleic Acids Res.* **41**, D991–D995 (2013).
11. Giovannini-Chami, L. *et al.* Distinct epithelial gene expression phenotypes in childhood respiratory allergy. *Eur. Respir. J.* **39**, 1197–1205 (2012).
12. Barrenas, F. *et al.* Highly interconnected genes in disease-specific networks are enriched for disease-associated polymorphisms. *Genome Biol.* **13**, R46 (2012).
13. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559 (2008).
14. Yu, G., Wang, L.G., Han, Y. & He, Q.Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* **16**, 284–287 (2012).
15. Gautier, L., Cope, L., Bolstad, B.M. & Irizarry, R.A. affy – analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics* **20**, 307–315 (2004).
16. Ritchie, M.E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).
17. Robin, X. *et al.* pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* **12**, 77 (2011).
18. Vercelli, D. Discovering susceptibility genes for asthma and allergy. *Nat. Rev. Immunol.* **8**, 169–182 (2008).
19. Pawankar, R., Bunnag, C., Khaltaev, N. & Bousquet, J. Allergic rhinitis and its impact on asthma in Asia Pacific and the ARIA update 2008. *World Allergy Organ. J.* **5**, S212–S217 (2012).
20. Shaoqing, Y. *et al.* Microarray analysis of differentially expressed microRNAs in allergic rhinitis. *Am. J. Rhinol. Allergy* **25**, e242–e246 (2011).
21. Dai, D.N. *et al.* Elevated expression of CST1 promotes breast cancer progression and predicts a poor prognosis. *J. Mol. Med. (Berl.)* **95**, 873–886 (2017).
22. Fukuoka, A. *et al.* Human cystatin SN is an endogenous protease inhibitor that prevents allergic rhinitis. *J. Allergy Clin. Immunol.* **143**, 1153–1162 (2019).
23. Imoto, Y. *et al.* Cystatin SN upregulation in patients with seasonal allergic rhinitis. *PLoS One* **8**, e67057 (2013).
24. Taha, M., Nezerwa, E. & Nam, H.J. The X-ray crystallographic structure of human EAT2 (SH2D1B). *Protein Pept. Lett.* **23**, 862–866 (2016).

25. Aldhamen, Y.A., Seregin, S.S., Aylsworth, C.F., Godbehere, S. & Amalfitano, A. Manipulation of EAT-2 expression promotes induction of multiple beneficial regulatory and effector functions of the human innate immune system as a novel immunomodulatory strategy. *Int. Immunol.* **26**, 291–303 (2014).
26. Nargis, T. & Chakrabarti, P. Significance of circulatory DPP4 activity in metabolic diseases. *IUBMB Life* **70**, 112–119 (2018).
27. Ghorpade, D.S. *et al.* Hepatocyte-secreted DPP4 in obesity promotes adipose inflammation and insulin resistance. *Nature* **555**, 673–677 (2018).
28. Tasic, T. *et al.* Dipeptidyl peptidase IV (DPP4) deficiency increases Th1-driven allergic contact dermatitis. *Clin. Exp. Allergy* **41**, 1098–1107 (2011).
29. Klemann, C., Wagner, L., Stephan, M. & von Horsten, S. Cut to the chase: a review of CD26/dipeptidyl peptidase-4's (DPP4) entanglement in the immune system. *Clin. Exp. Immunol.* **185**, 1–21 (2016).
30. de Morais, R.M., Sobrinho, A.B., de Souza Silva, C.M., de Oliveira, J.R., da Silva, I.C.R. & de Toledo Nobrega, O. The role of the NIS (SLC5A5) gene in papillary thyroid cancer: a systematic review. *Int. J. Endocrinol.*. **2018**, 9128754 (2018).
31. Guerrieri, F. *et al.* The sodium/iodide symporter NIS is a transcriptional target of the p53-family members in liver cancer cells. *Cell Death Dis.* **4**, e807 (2013).
32. Padiernos, R.B. & Mingala, C.N. Molecular comparison of Slc11a1 and Slc11a2 genes of swamp- and riverine-type water buffaloes. *Int. J. Immunogenet.* **43**, 171–179 (2016).
33. Meng, H., Liu, Y. & Lai, L. Diverse ways of perturbing the human arachidonic acid metabolic network to control inflammation. *Acc. Chem. Res.* **48**, 2242–2250 (2015).
34. Gunawardhana, L.P., Gibson, P.G., Simpson, J.L., Benton, M.C., Lea, R.A. & Baines, K.J. Characteristic DNA methylation profiles in peripheral blood monocytes are associated with inflammatory phenotypes of asthma. *Epigenetics* **9**, 1302–1316 (2014).
35. Lee, S.N., Ahn, J.S., Lee, S.G., Lee, H.S., Choi, A.M.K. & Yoon, J.H. Integrins $\alpha_v\beta_5$ and $\alpha_v\beta_6$ mediate IL-4-induced collective migration in human airway epithelial cells. *Am. J. Respir. Cell Mol. Biol.* **60**, 420–433 (2019).
36. Shi, Z. *et al.* Exploring the key genes and pathways in enchondromas using a gene expression microarray. *Oncotarget* **8**, 43967–43977 (2017).
37. Xuan, L. *et al.* MicroRNAs regulating mucin type O-glycan biosynthesis and transforming growth factor beta signaling pathways in nasal mucosa of patients with chronic rhinosinusitis with nasal polyps in Northern China. *Int. Forum Allergy Rhinol.* **9**, 106–113 (2019).

© 2019 The Authors. *Clinical and Translational Science* published by Wiley Periodicals, Inc. on behalf of the American Society for Clinical Pharmacology and Therapeutics. This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

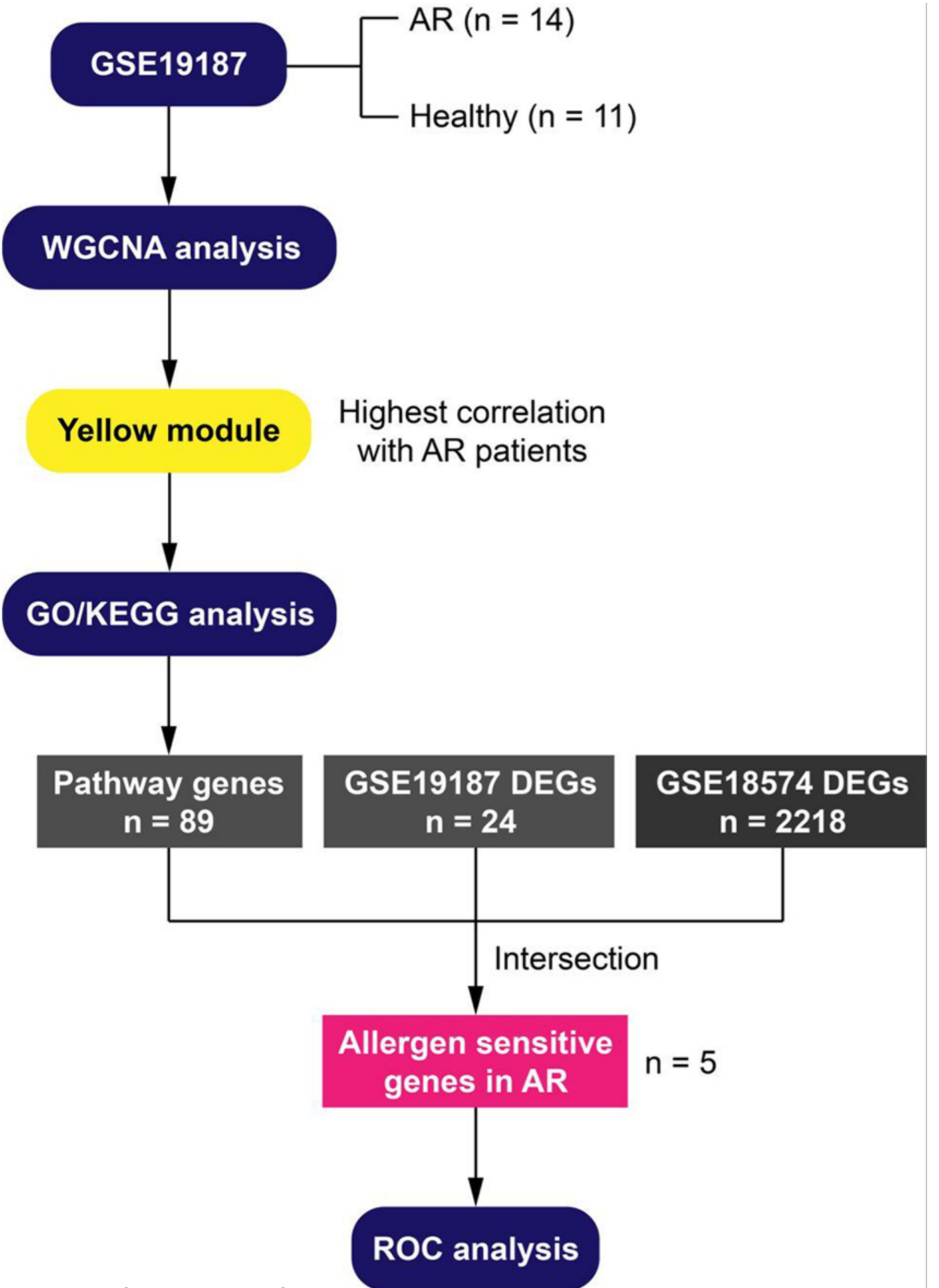

Figure S1 The flow diagram of our protocol