

RESEARCH

Open Access



CTPAD: an interactive web application for comprehensive transcriptomic profiling in allergic diseases

Suizi Zhou^{1,2†}, Wanqiao Huang^{3†}, Yitong Liu², Peng Luo³, Anqi Lin^{3*}, Hong Yang^{3*} and Qianhui Qiu^{1,2*}

Abstract

Background Allergic diseases are systemic chronic inflammatory diseases associated with multiorgan damage and complex pathogenesis. Several studies have revealed the association of gene expression abnormalities with the development of allergic diseases, but the biomedical field still lacks a public platform for comprehensive analysis and visualization of transcriptomic data of allergic diseases.

Objective The aim of the study is to provide a comprehensive web tool for multiple analysis in allergic diseases.

Methods We retrieved and downloaded human and mouse gene expression profile data associated with allergic diseases from the Gene Expression Omnibus (GEO) database and standardized the data uniformly. We used gene sets obtained from the MSigDB database for pathway enrichment analysis and multiple immune infiltration algorithms for the estimation of immune cell proportion. The basic construction of the web pages was based on the Shiny framework. Additionally, more convenient features were added to the server to improve the efficiency of the web pages, such as jQuery plugins and a comment box to collect user feedback.

Results We developed CTPAD, an interactive R Shiny application that integrates public databases and multiple algorithms to explore allergic disease-related datasets and implement rich transcriptomic visualization capabilities, including gene expression analysis, pathway enrichment analysis, immune infiltration analysis, correlation analysis, and single-cell RNA sequencing analysis. All functional modules offer customization options and can be downloaded in PDF format with high-resolution images.

Conclusions CTPAD largely facilitates the work of researchers without bioinformatics background to enable them to better explore the transcriptomic features associated with allergic diseases. CTPAD is available at <https://smuonco.shinyapps.io/CTPAD/>.

Keywords Allergic diseases, Transcriptomic profiling, Visualization, Gene, Bioinformatics

[†]Suizi Zhou and Wanqiao Huang have contributed equally to this work.

*Correspondence:

Anqi Lin

smulinanqi0206@i.smu.edu.cn

Hong Yang

smuyanghong@i.smu.edu.cn

Qianhui Qiu

qiuqianhui@gdph.org.cn

Full list of author information is available at the end of the article



Introduction

Allergic diseases represent a group of conditions caused by hypersensitivity of the immune system to allergens present in the environment [1]. These diseases, which are mediated by anti-immunoglobulin E, include bronchial asthma (AS), food allergy (FA), drug allergy (DA), atopic dermatitis (AD), allergic rhinitis (AR), conjunctivitis and chronic rhinosinusitis with or without nasal polyposis [2–4]. Over the past few decades, with the acceleration of industrialization, increases in environmental pollution, changes in lifestyle and dietary structure, and continuous increases in allergen exposure, allergy diseases have come to affect 30–40% of the global population and impose a major public health burden worldwide [5–8], with an estimated 300 million patients with AS, 400 million patients with AR, 200 million to 250 million patients with FA, 150 million patients with DA and numerous patients with allergic conjunctivitis, angioneurotic oedema, urticaria, eczema, eosinophilic diseases, insect allergy, and anaphylactic shock. Concomitantly, allergy is listed by the World Health Organization as one of the 6 major chronic diseases of the twenty-first century [9].

High-throughput sequencing technologies can help reveal genetic and epigenetic influences on allergic diseases and thus reveal potential allergenic genes. A meta-analysis of allergic diseases showed that thousands of genes are associated with allergic diseases, and some of them are also associated with processes such as immune inflammation, cytokines, and viral infections [10]. Moreover, high-throughput sequencing technologies can be used to discover new biomarkers to enable more accurate diagnoses of allergic diseases. One study identified some genes associated with respiratory diseases, among which HLA-DQ and SCGB1A1 were the most highly associated with asthma; thus, these genes could be used as new disease markers to diagnose asthma and other respiratory diseases [11].

An increasing number of studies based on transcriptomic datasets have revealed the association between genetic susceptibility and allergic diseases [12–15]. However, most transcriptomic databases only provide processed data within their respective studies; no data integration in these databases has been carried out in an app, and no batch-corrected values are provided. To obtain mechanistic insights into how genes may modify biological pathways relevant to a given trait under consideration, public gene expression data from resources such as the Gene Expression Omnibus (GEO) are a primary resource for answering the above questions. Many experimental researchers do not have the expertise or dedicated computational resources necessary to obtain and integrate gene expression microarray, RNA-Sequencing results, and so on. Even researchers who do have such

resources may repeat similar analytical tasks every time a new association study is performed.

Integrated atlases for select species, tissues, or diseases related to allergies are highly useful as consensus reference maps and for enhancing downstream analyses. For instance, in oncology research, interactive databases such as cBioPortal [16], GEPIA2 [17], OncoDB [18], and UCSCXenaShiny [19] have become invaluable tools for analyzing cancer multiomic data. Since 2013, approximately 16,000 publications have cited these resources, leveraging their capabilities to validate research findings. These web applications facilitate a wide range of analyses, from routine transcriptomic data analysis—including differential gene expression, pathway enrichment, correlation, and Kaplan–Meier survival analyses—to more advanced techniques such as single-cell RNA sequencing, methylomic, and genomic data analyses. Moreover, these platforms excel in delivering high-quality visualizations, making them essential for both basic and translational cancer research. However, in the realm of allergic diseases, a comparable comprehensive platform is currently absent. Therefore, there is an urgent imperative to develop a robust platform dedicated to exploring gene-disease associations specifically tailored for immune alterations in allergic disease.

To fill this gap, we aimed to collect data from the GEO database on allergic diseases. Differentially expressed genes, pathway enrichment, immune infiltration, the correlation between genes and pathways, and single-cell RNA sequencing data analysis were used to explore the relationship between genes and disease. Hopefully, our study provides valuable data resources and a freely available platform to investigate gene expression and immune trajectory across allergic diseases that can contribute to elucidating the immune-related mechanisms of allergic diseases and that can strengthen the clinical management of these diseases.

Materials and methods

Data preprocessing

We collected and downloaded 46 transcriptomic datasets (both single cell sequencing and bulk sequencing) related to allergic diseases from the GEO database [20] (Table S1). We performed a secondary manual screening and review of the included samples. We included only baseline data for replicate trials at different time points or drug concentrations. All samples were collected from untreated patients. Cell lines and data that involved flow cytometry sorting were excluded. A total of 1889 tissue samples and peripheral blood samples were eventually included in this study. The included transcriptome data were annotated with gene symbols, normalized, and log₂ transformed. Among them, microarray data

were normalized using the quantile method from the limma package (version 3.48.3) [21]. For high-throughput sequencing data with raw count data, normalization was performed using the TMM method from the edgeR package (version 3.34.0) [22–24]. For high-throughput sequencing datasets lacking raw count data, NCBI-generated RNA-seq count data were utilized and normalized using the TMM-edgeR method, as previously described.

Data analysis

In the differential expression analysis, the limma package [21] was used for differential analysis of the expression array for gene microarray datasets, while the edgeR package [22–24] was used for differential analysis of high-throughput sequencing datasets.

For the enrichment analysis, we first collected 13,661 gene sets from the MSigDB database [25] for humans and 12,297 gene sets for mice (Table S2). Among them, the human gene set includes 50 hallmark gene sets, 3050 canonical pathway (CP) gene sets, and 10,561 Gene Ontology (GO) gene sets; the mouse gene set includes 50 MH gene sets, 1687 CP gene sets and 10,560 GO gene sets. For all transcriptome data, we performed gene set enrichment analysis (GSEA) pathway enrichment analysis of the differential analysis results using the clusterProfiler package (version 4.0.5) [26]. In addition, the normalized transcriptome data were subjected to single-sample gene set enrichment analysis (ssGSEA) with the GSVA package (version 1.40.1) [27].

For immune infiltration analysis, we used the immune-deconv package (version 2.1.3) [28, 29] to assess the level of immune cell infiltration in samples. The analysis utilized several algorithms, including CIBERSORT [30], MCPCounter (human) [31], mMCPCounter (mouse) [32], quanTIseq [33], xCell [34] and EPIC [35]. Significance P values were calculated by the Wilcoxon rank sum test.

For correlation analysis, gene-to-gene, gene-to-pathway, and pathway-to-pathway correlations were performed using normalized transcriptomic data with ssGSEA scores. Pearson and Spearman correlation analyses were used to calculate correlations between variables.

For single-cell RNA sequencing analysis, 8 single-cell RNA-seq datasets were analyzed using the Seurat package (version 4.4.0) [36]. Cells with UMI counts greater than 2500 or less than 200, and mitochondrial percentages exceeding 5% were filtered out. The remaining data were log-transformed using the “NormalizeData” function and assessed for variability using the “FindVariableFeatures” function with the “vst” method, identifying 2000 highly variable features per dataset. Principal component analysis (PCA) clustering was conducted using the “RunPCA” function, followed by non-linear

dimensionality reduction using Uniform Manifold Approximation and Projection (UMAP) for improved data visualization. Cellular markers were identified using the “FindAllMarkers” function, and differentially expressed genes (DEGs) were filtered based on log₂ fold change (> 1). Cell types were manually annotated based on DEGs expressed in different cells.

Data visualization

For visualization of the analysis results, we used the ggplot2 package (version 3.5.0) [37] to generate volcano plots for differential analysis, stacked histograms for immune infiltration, box plots for immune infiltration, and scatter diagrams for correlation analysis. Heatmap visualization for the differential expression analysis, enrichment analysis, and immune infiltration analysis was implemented with the ComplexHeatmap package (version 2.8.0) [38, 39]. Dot plots and ridge plots for enrichment analysis were generated with the enrichplot package (version 1.12.2) [40], and GSEA maps were generated with the GeseaVis package (version 0.1.0) [41]. Correlation heatmaps were plotted using the corrplot package (version 0.92) [42]. Single-cell sequencing data were visualized using the “FeaturePlot” function from the Seurat package.

Shiny web application construction

CTPAD was built with the shiny package (version 1.8.1.1) [43]. No user login is required to access any of the functional modules of the CTPAD. High-resolution PDF image download capability is provided for all analytical visualization results of CTPAD. The data tables corresponding to the analysis results are generated by DT package (version 0.33) [44], allowing users to search, sort, and download the data (in CSV format). CTPAD is available at <https://smuonco.shinyapps.io/CTPAD/> (alternative URL: <http://robinl-lab.com/CTPAD/>), with the corresponding code accessible in the public GitHub repository (<https://github.com/ZJYY-ONCOLOGY/CTPAD>).

Nasal sample collection

This study was approved by the Institutional Review Boards of Zhujiang Hospital, Southern Medical University, China. All participants provided informed consent in accordance with the Declaration of Helsinki. Biopsy specimens of the inferior turbinate (IT) were obtained from control individuals and AR patients with septal deviation during septal plastic surgery. AR was diagnosed according to the Initiative on Allergic Rhinitis and Its Impact on Asthma guidelines [45], based on either the skin prick test (Alutard, ALK-Abellórd, Denmark) or serum total immunoglobulin E detected using an allergy screen test (LG Chem, South Korea). All recruited AR patients were

free of rhinosinusitis, lower respiratory tract infections, and self-reported or physician-diagnosed asthma and smoking.

After excision, all samples were washed with saline solution to remove surface bloodstains and processed within 15 min as follows: Samples designated for histological analysis were fixed in 4% paraformaldehyde tissue fixative (Biosharp, BL539A) for 24 h. Following fixation, samples were embedded in paraffin and continuously sectioned at 5 μ m thickness. The nasal tissue sections were then placed on a slide warmer and baked at 60 °C for over 1 h for subsequent analysis.

Immunohistochemistry (IHC)

Paraffin sections were deparaffinized by immersing them in xylene for 20 min each, followed by a series of ethanol and distilled water washes. Antigen retrieval was performed by boiling the sections in antigen retrieval solution, followed by cooling to room temperature and washing with phosphate-buffered saline (PBS). To block endogenous peroxidase activity, tissue sections were circled with an IHC hydrophobic barrier pen, treated with hydrogen peroxide blocking reagent, incubated for 20 min, and washed with PBS.

To identify and localize the protein expression of MUC2, SLC7A1 (CAT-1), and BCL2L15, IHC staining was utilized. Tissue sections were incubated with primary antibodies overnight at 4 °C, then washed with PBS. This was followed by secondary antibody incubation using horseradish peroxidase (HRP) polymer for 20 min at room temperature and another PBS wash. The 3,3'-diaminobenzidine (DAB) working solution was prepared by mixing DAB buffer with DAB solution, and DAB staining was performed for 1–5 min, followed by rinsing with running water. Sections were then counterstained with hematoxylin, rinsed, and dehydrated through a series of ethanol and xylene washes. Finally, the sections were mounted with neutral balsam.

The primary antibodies used were as follows: MUC2 polyclonal antibody (1:2000, 27675-1-AP, Proteintech), SLC7A1/CAT-1 polyclonal antibody (1:200, 14195-1-AP, Proteintech), and BCL2L15 polyclonal antibody (1:250, 23975-1-AP, Proteintech).

IHC image quantification analysis

IHC images were captured using a Leica DM6 microscope and analyzed with ImageJ software (version 1.54j). The study included both control and AR groups, with three IHC images of the inferior nasal concha mucosa randomly selected from each group. These images consistently demonstrated the integrity of the nasal mucosal epithelium.

Quantitative analysis involved selecting three random fields of view from each image to measure the percentage of positively stained area. ImageJ software processed the images, specifically using color deconvolution to analyze DAB-stained regions. A uniform measurement threshold ensured consistency and comparability across all images. The percentage of positive area was calculated by dividing the stained area by the total area of a fixed rectangular frame. Statistical significance was assessed using the Wilcoxon rank sum test.

Results

Overview

CTPAD is a shiny web tool for the study of allergic diseases. The web tool incorporates 46 allergic disease-related datasets (both single cell sequencing and bulk sequencing) from the GEO database, containing 2 species, 9 diseases, and 1,889 samples (Fig. 1). The CTPAD home page provides a flowchart for the web tool and sample visualizations of each functional module for a quick overview of the main functions of the web page. All visualizations are provided with a download button for easy access to high-resolution vector images. The Data page integrates detailed descriptions of the datasets included in CTPAD. Users can obtain detailed answers to common questions they may encounter when using CTPAD on the about page or can communicate with the authors according to the contact information provided.

Differential expression analysis module

The Differential Expression Analysis module allows users to access information on the differential expression of genes in the dataset and to explore statistically significant up- and downregulated genes. For gene differential expression analysis results, two visualizations are provided: the volcano plot (Fig. 2A) and the heatmap (Fig. 2B). In the Volcano Plot tab, users can first select a dataset of interest by searching for disease type or dataset name. The default screening criteria for differentially expressed genes were a p value less than 0.05 and $|\log_2(\text{FoldChange})|$ over 2. Additionally, the names of 10 significantly upregulated and 10 significantly downregulated differentially expressed genes are clearly labelled on the volcano plot. The above threshold criteria and gene name annotations can be freely adjusted according to user requirements. Of note, to enhance advanced visualization capabilities and better meet user needs, CTPAD supports customizable color patterns and gene selection on volcano plots. In the Heatmap tab, after determining the dataset of interest and differential gene screening criteria, users can choose between two heatmap visualization options: plotting the top 10/top 5 up- and

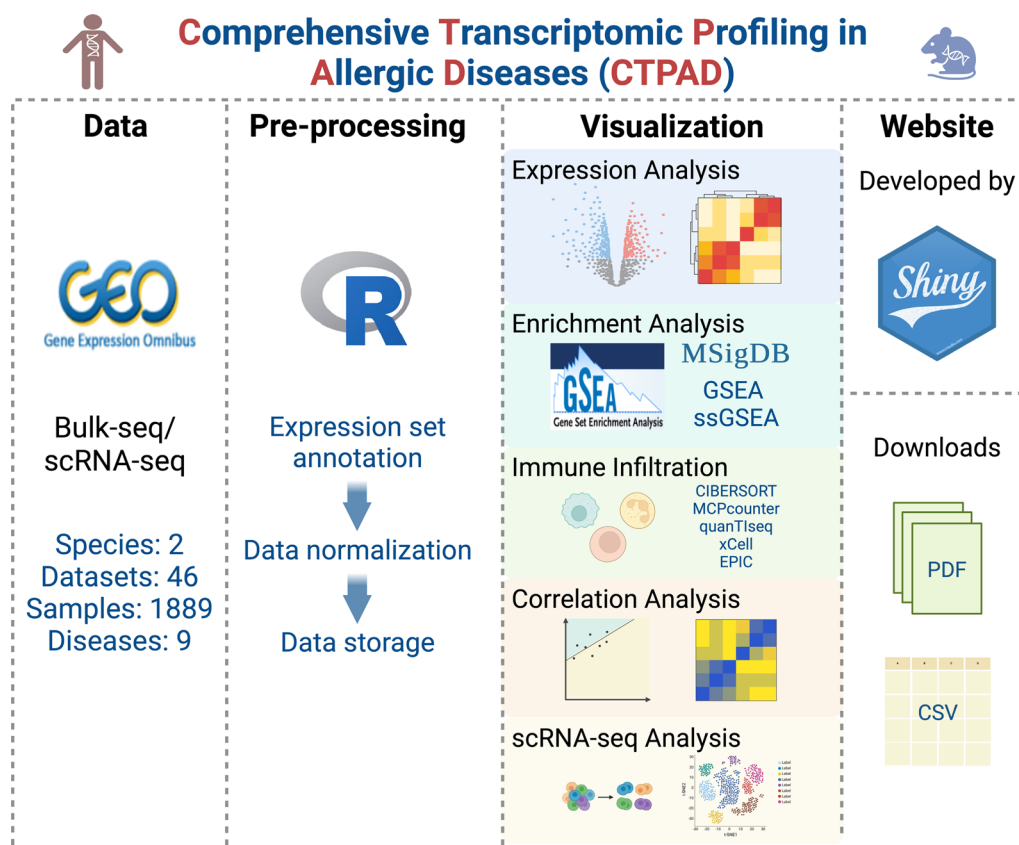


Fig. 1 Workflow diagram presenting data collection, preprocessing, and web tool construction for the CTPAD web tool (Created with BioRender.com)

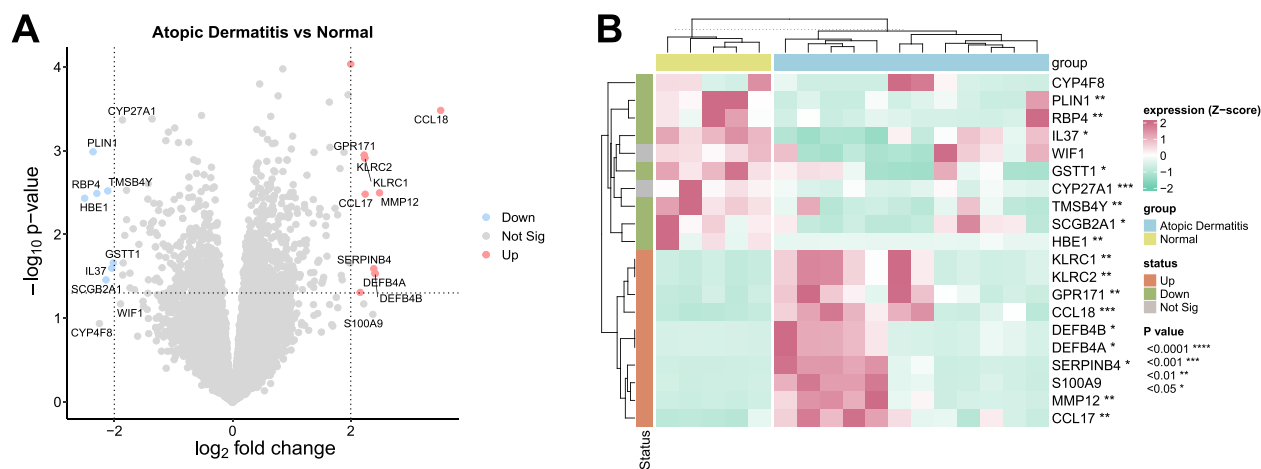


Fig. 2 Volcano plot and heatmap visualization results based on DEG analysis results in the Expression Analysis module. **A** Volcano plot showing the names of significantly up- and downregulated differentially expressed genes in the GSE5667 dataset. **B** Heatmap depicting gene expression between atopic dermatitis (AD) and controls, with red representing upregulation and green representing downregulation, where p values were calculated by the t test built into the limma package. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$

downregulated differentially expressed genes (default) or searching and selecting the gene symbol of interest for plotting.

Enrichment analysis module

The enrichment analysis module provides two common pathway enrichment analysis algorithms, GSEA

and ssGSEA, for users to explore the enrichment status of different pathways in the target dataset. Users can explore nearly 25,000 human and mouse pathways from the MsigDB database. For the GSEA results, three visualizations are provided: point plots (Fig. 3A), ridge plots (Fig. 3B), and GSEA plots (Fig. 3C). Notably, CTPAD provides rich visualization customization functions. For point plots, users can adjust the number of gene sets displayed for visualization. In GSEA plots, users can choose to annotate the visualization results with the gene symbols that contribute most to the enrichment score and

can customize the number of gene annotations. For the ssGSEA results, users can select the heatmap (Fig. 3D) for visualization. The normalized ssGSEA enrichment scores, detailed grouping information, and significance of pathway enrichment differences are presented on the heatmap. Similarly, the number of visualized pathways can be adjusted by the user as well.

Immune infiltration module

The Immune Infiltration module provides users with five commonly used immune infiltration algorithms to

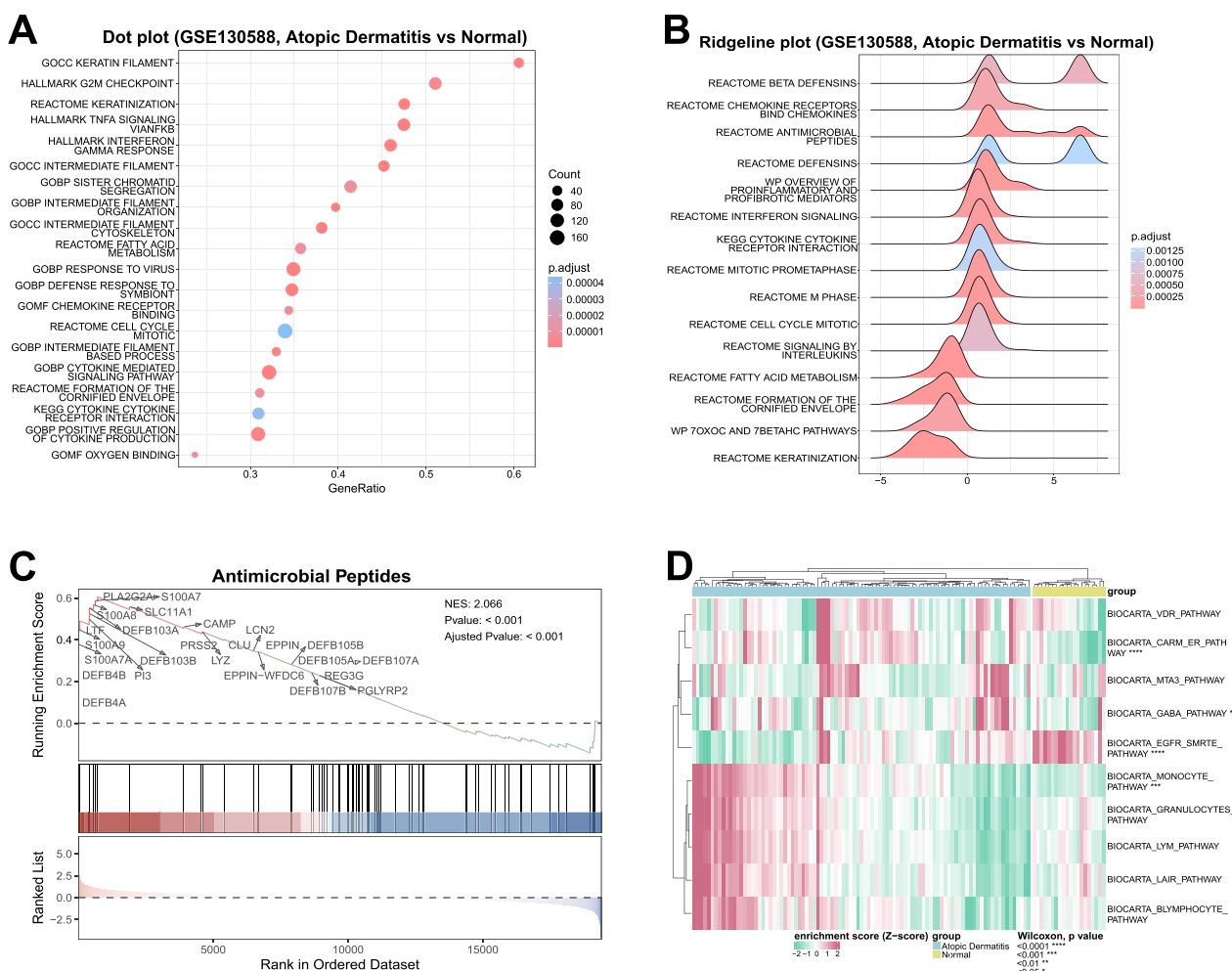


Fig. 3 Multiple visualization results of pathway enrichment analysis in the Enrichment Analysis module. **A** GSEA enrichment results between the atopic dermatitis group and the normal group. The scatter diagram shows the top 20 gene sets sorted according to the p-adjusted value. **B** Expression distribution of core enriched genes in the set of CP genes enriched by GSEA between the atopic dermatitis group and the normal group. The x-axis is log2-fold of the change in expression of core enriched genes in the enrichment pathway, > 0 indicates upregulated expression, and < 0 indicates downregulated expression. **C** GSEA plot showing the enrichment of the REACTOME_ANTIMICROBIAL_PEPTIDES pathway in the atopic dermatitis group compared to the normal group. The top 25 genes in the gene set that contributed most to the GSEA enrichment score are labelled in the figure. **D** Differences in ssGSEA enrichment scores between the atopic dermatitis group and the normal group; red represents upregulation, and green represents downregulation. p values were calculated by the Wilcoxon test. *p < 0.05, **p < 0.01, ***p < 0.001, ****p < 0.0001

explore the infiltration of immune cells in different subgroups of the target dataset. Three visualization methods, stacked histogram, heatmap, and box plot, are provided for users to choose from according to their research needs. Among them, the stacked bar chart provides an overview of the proportion of different immune cell infiltrations between the disease and control groups. If users want to visualize the significant differences in immune infiltration under different groupings, they can choose a heatmap and boxplot to visualize the results. Taking the GSE148240 dataset as an example, Fig. 4A–C shows the corresponding three visualization result plots.

Correlation analysis module

The correlation analysis module allows users to explore correlations between genes, between pathways, and between genes and pathways. To meet the need for the number of variables in different studies, scatter diagrams and correlation heatmaps are provided. Among these visualizations, the scatter diagram supports studies exploring two variables (gene and gene, gene and pathway, pathway and pathway), while the correlation heatmap can support correlation studies among up to 20 genes. Both visualizations provide 3 study groups (disease group only, control group only, disease group vs. control group) and 2 statistical methods of correlation analysis, namely Spearman correlation coefficient and Pearson correlation coefficient) for users to flexibly choose from. Users are free to search for gene or pathway names of interest in the target dataset according to the purpose of the study. For example, in the GSE150910 dataset, MUC5B expression was previously found to be associated with epithelial cell development in patients with chronic allergic pneumonia by weighted gene coexpression network analysis [46]. We analysed and visualized the correlation between MUC5B expression and the GOBP_EPIDERMIS_DEVELOPMENT pathway in the disease group of this dataset by scatter diagram (Fig. 5A), and this result verified the positive correlation between the two. For the gene correlation heatmap, the user is free to adjust the number of genes visualized, and the correlation coefficients and their significance are visually and clearly labelled on the graph (Fig. 5B).

Single-cell RNA sequencing analysis module

The single-cell RNA sequencing analysis module in CTPAD enables researchers to explore intercellular heterogeneity, identify molecular markers within the transcriptome, and assess pathway activation statuses at the cellular subpopulation level. This technique offers a significant advancement over bulk sequencing, which averages gene expression across the entire tissue sample, thereby masking the rich diversity of cellular

subpopulations and the behaviors of individual cells. In CTPAD, users have access to powerful visualization tools such as cluster plots, feature plots and heatmap, which are instrumental in analyzing and interpreting embedded scRNA-seq data. Cluster plots (Fig. 6A, B) are essential for visualizing the distinct cellular subpopulations within a sample. CTPAD employs UMAP for dimensional reduction, where distinct colored clusters represent different cell types or states, grouped based on their gene expression profiles. Feature plots and Heatmap (Fig. 6C–E) allow users to investigate the expression levels of specific genes or pathways across different cell populations by utilizing DEGs and ssGSEA scores.

Validation of CTPAD findings via IHC: exploring DEGs associated with AR

Using CTPAD, we conducted a comprehensive DEG analysis across multiple datasets to investigate gene biomarkers associated with AR. We included eight datasets encompassing human nasal biopsies from both control and AR patients (GSE19187, GSE43523, GSE44037, GSE46171, GSE51392, GSE118243, GSE119136, and GSE206149). DEGs were defined as genes with a p-value less than 0.05 and an absolute $|\log_2(\text{FoldChange})|$ greater than 1. Our analysis revealed that in 5 out of the 8 human AR datasets, three genes (BCL2L15, MUC2, SLC7A1) were significantly upregulated in the disease group compared to the normal control group. This observation was further validated using IHC on samples from our local cohort. Consistent with the CTPAD bioinformatic results, BCL2L15 and MUC2 exhibited significantly enhanced protein expression in the inferior nasal concha mucosa of the AR group (Fig. 7). Therefore, we propose that BCL2L15 and MUC2 are potential novel biomarkers for allergic rhinitis.

Discussion

CTPAD is the first comprehensive web tool that has been developed for the integration of multiple analysis tools and interactive analysis in allergic diseases. The data in CTPAD include gene expression data from 46 datasets, from both single-cell RNA sequencing data and bulk RNA sequencing data, from 1889 samples, from mice and humans, and from blood and tissue in 9 common allergic diseases. CTPAD is a time-saving, free, and intuitive tool for tapping the full potential of publicly available transcriptomics data, which enables biologists and clinicians without any programming experience to visualize allergy-related gene expression and immune profiles and to perform a diverse range of data analyses.

CTPAD represents a pioneering web tool designed to integrate and analyze comprehensive gene expression data across diverse allergic diseases. On the one hand,

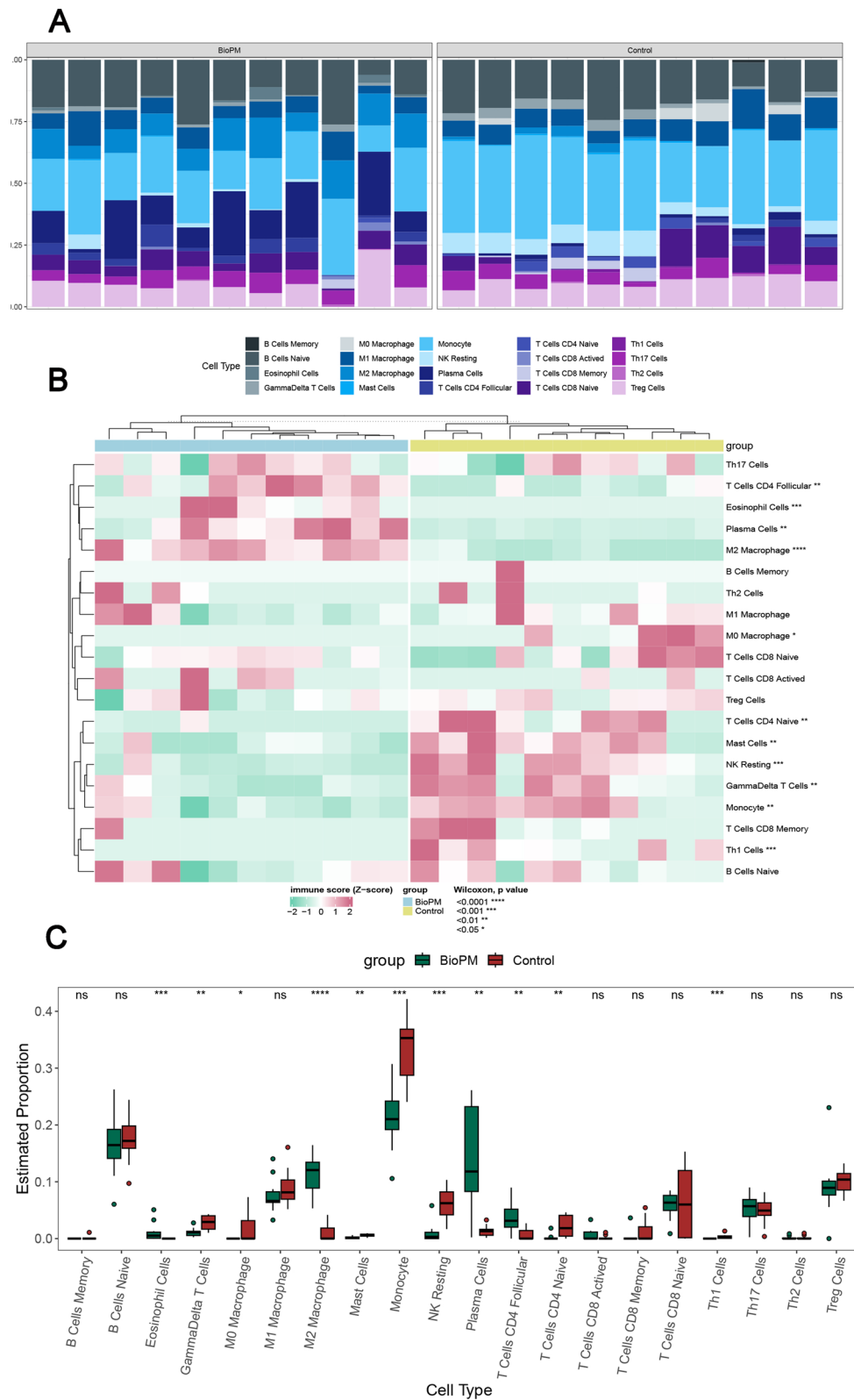


Fig. 4 Multiple visualization results for immune infiltration analysis in the Immune Infiltration module. **A** Overview of the different immune cell infiltration proportions between the BioPM-exposed and control groups. **B** Heatmap showing the difference in immune infiltration scores between the BioPM-exposed and control groups. **C** Boxplot visualizing the difference in the difference in the percentage of immune cell infiltration. p values were calculated by the Wilcoxon test. BioPM: biological particulate matter. *p < 0.05, **p < 0.01, ***p < 0.001, ****p < 0.0001

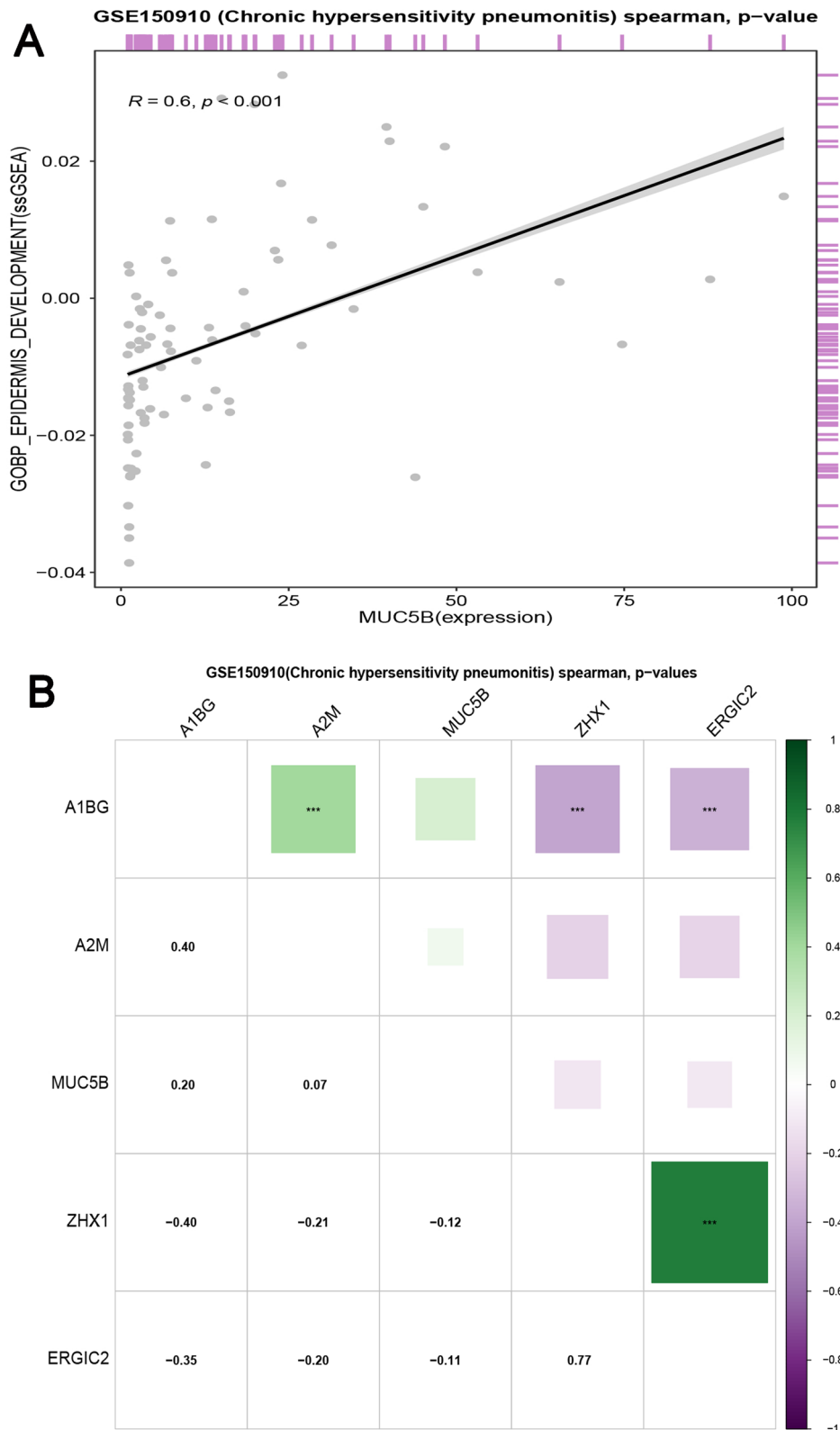


Fig. 5 Two visualizations of correlation analysis in the Correlation Analysis module. **A** Scatter diagram showing a positive correlation between MUC5B mRNA expression and GOBP_EPIDERMIS_DEVELOPMENT pathway ssGSEA score ($R = 0.6, p < 0.001$). **B** Correlation heatmap showing the results of correlation analysis between the 5 genes; green represents a positive correlation, and purple represents a negative correlation. R is the Spearman correlation coefficient. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$

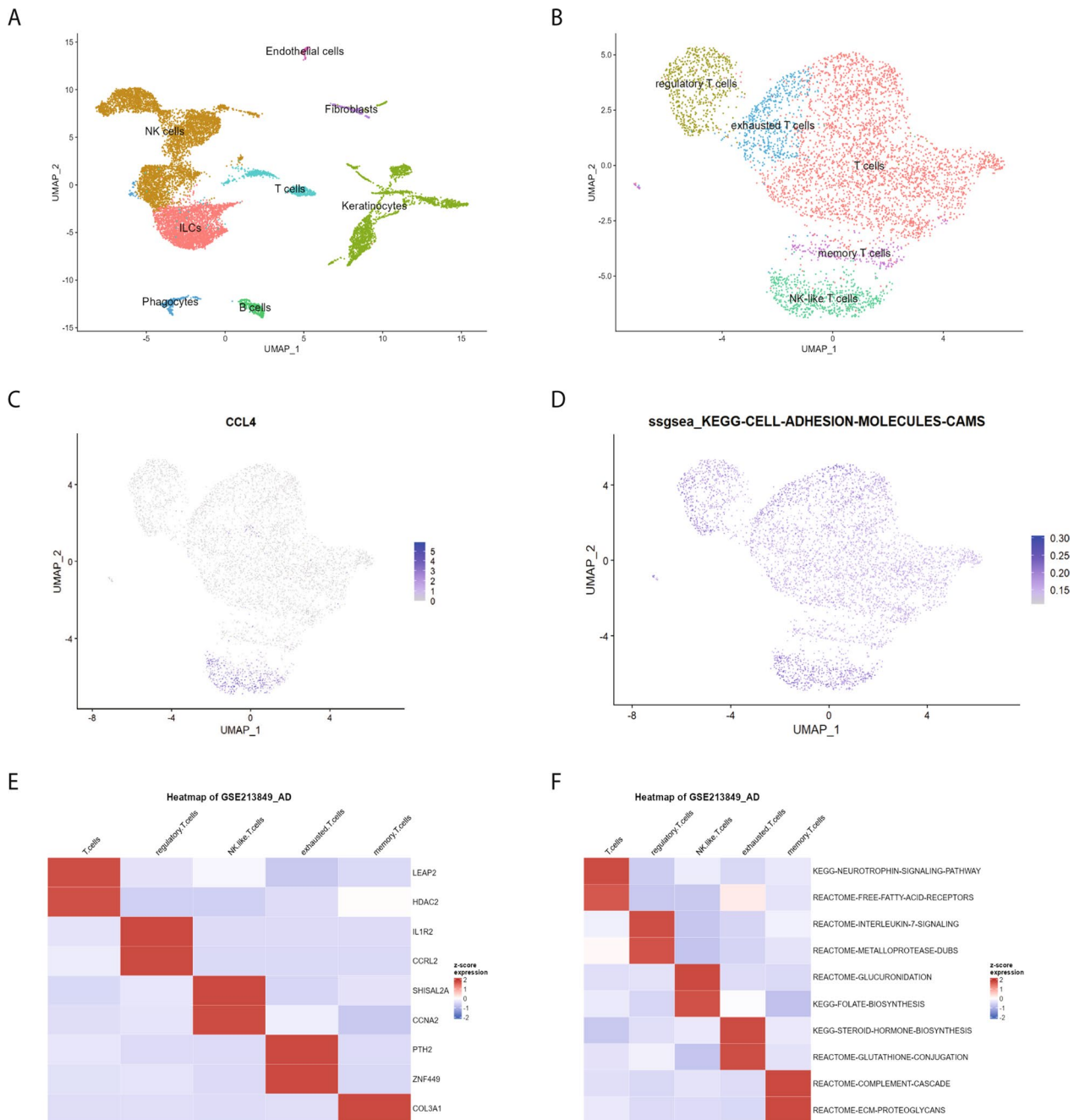


Fig. 6 Visualizations of Single-cell RNA Sequencing Analysis Module. **A, B** Cluster plot displaying distinct cell subtype clusters following UMAP dimensionality reduction. Each point represents an individual cell colored according to its assigned cluster identity. **C, D** Feature plot illustrating the expression patterns of genes of interest and pathway activation profiles using the ssGSEA algorithm. **E, F** Heatmap showing the expression levels of genes of interest and pathway activation profiles using the ssGSEA algorithm. ssGSEA: single sample Gene Set Enrichment Analysis; UMAP: Uniform Manifold Approximation and Projection

we offer users diverse features in terms of disease-related allergy, analysis methods, visualization, and customization. On the other hand, we identified the functionally distinct cell types that comprise the immune response, assessing immune infiltration and determining whether

differences in the composition of immune infiltration can improve the development of novel immunotherapeutic drugs to target these cells. Most importantly, the reliability of CTPAD results is further validated through bench experiments. In CTPAD, DEG analysis reveals

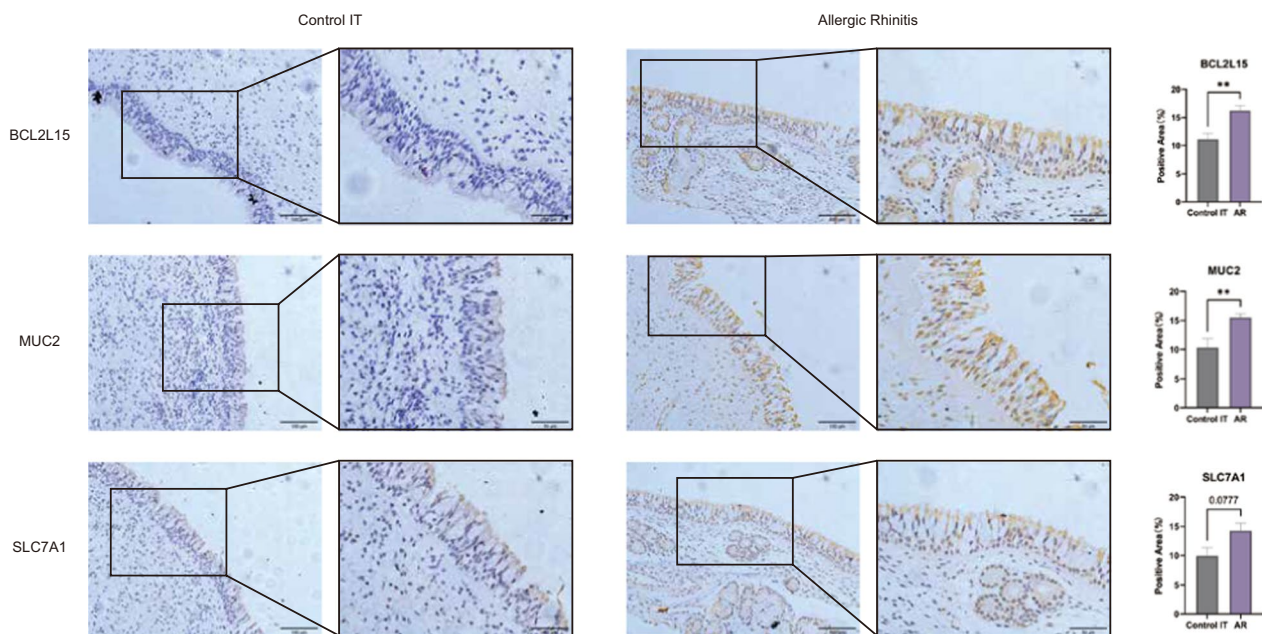


Fig. 7 Validation of allergic rhinitis-related differentially expressed genes via IHC. IHC images and corresponding quantification demonstrate the protein expression levels of BCL2L15, MUC2, and SLC7A1 in the inferior nasal concha mucosa of control and AR groups. BCL2L15 and MUC2 show significant upregulation in the AR group ($p < 0.01$), while SLC7A1 exhibits elevated expression without reaching statistical significance ($p = 0.0777$). The percentage of positive area was calculated by dividing the stained area by the total area of a fixed rectangular frame. Significance was determined using the Wilcoxon rank sum test. AR: Allergic Rhinitis; IHC: Immunohistochemistry. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$

that three genes (BCL2L15, MUC2, SLC7A1) are significantly upregulated at the mRNA level between the control group and the AR group. This observation is corroborated at the protein level by IHC staining, which shows increased protein expression of BCL2L15 and MUC2 in the inferior nasal concha mucosa of the AR group. This proposition is partially supported by previous publications indicating the significant roles of these genes in the immune system and inflammation [47–49]. Notably, MUC2 encodes a mucin protein, which is a high molecular weight glycoprotein and a major component of mucus [50]. Excessive secretion of mucus is an important hallmark of many allergic diseases, ranging from AR to asthma. Hence, CTPAD holds significant promise for identifying novel biomarkers for allergic diseases and robustly validating these findings through bioinformatic analysis, thereby bridging the gap between computational predictions and experimental validation.

Nevertheless, our study is subject to several limitations. Firstly, we acknowledge potential biases or challenges associated with data retrieval and preprocessing. Despite conducting thorough manual screening in four rounds and employing double-check procedures, the possibility of missing datasets remains. Also, our inclusion criteria are constrained to datasets available until May 2024, thereby restricting our analysis to a specific timeframe. To mitigate this issue, we have implemented

an alternative module on the CTPAD website, enabling users to submit dataset requests via a comment box. Another limitation pertains to the current inability of our platform to perform integrated analysis of multiple datasets or to accommodate user-uploaded data. To counteract these limitations, we are committed to regularly updating the CTPAD website every 3–4 months with newly acquired datasets and functions, ensuring ongoing relevance and comprehensiveness. We intend for CTPAD to be a comprehensive and high-quality repository for processed gene data.

To better serve the allergic disease research community, we will not only continuously update the allergic disease-related datasets as new studies are published but also develop new analytical features for further exploration of the available big genomic data. Our next plan is to obtain public multiomics data on allergic diseases and build an enhanced database based on comprehensive genome-level data for the effective visualization and analysis of all human genes in the future. As the functions of more genes are revealed in specific diseases, CTPAD will become a useful platform for both bench scientists and computational biologists and will contribute to clinical and translational studies.

Abbreviations

AD	Atopic dermatitis
AR	Allergic rhinitis

AS	Bronchial asthma
CP	Canonical pathway
CTPAD	An interactive web application for Comprehensive Transcriptomic Profiling in Allergic Diseases.
DA	Drug allergy
DAB	3,3'-Diaminobenzidine
DEG	Differentially expressed genes
FA	Food allergy
GEO	Gene expression omnibus
GO	Gene ontology
GSEA	Gene set enrichment analysis
IHC	Immunohistochemistry
IT	Inferior turbinate
PBS	Phosphate-buffered saline
PCA	Principal component analysis
ssGSEA	Single-sample Gene Set Enrichment Analysis
UMAP	Uniform Manifold Approximation and Projection

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12967-024-05459-2>.

Supplementary Material 1: Table S1: Overview of allergic disease-related datasets included in CTPAD.

Supplementary Material 2: Table S2: Gene sets applied to pathway enrichment analysis (downloaded from MSigDB database).

Acknowledgements

Not applicable.

Author contributions

Suizi Zhou designed the project and wrote the main manuscript. Wanqiao Huang and Hong Yang did the bioinformatics analysis, revised the manuscript, and prepared figures. Yitong Liu collected the data and performed the immunohistochemistry staining. Peng Luo and Anqi Lin supervised the project, analyzed the data and revised the manuscript. Qianhui Qiu led and funded the project.

Funding

This study was supported by Grants from National Nature Science Foundation of China No. 82171104 and 82371113 (to Qianhui Qiu) and National Postdoctoral Researcher Support Program No. GZC20230576 (to Suizi Zhou).

Availability of data and materials

All datasets used in this study were downloaded from the GEO database (www.ncbi.nlm.nih.gov/geo). CTPAD is available at <https://smuonco.shinyapps.io/CTPAD/> (alternative URL: <http://robinl-lab.com/CTPAD/>), with the corresponding code accessible in the public GitHub repository (<https://github.com/ZJYY-ONCOLOGY/CTPAD>).

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Approved.

Competing interests

The authors declare no potential conflicts of interest.

Author details

¹Guangdong Cardiovascular Institute, Guangdong Provincial People's Hospital, Guangdong Academy of Medical Sciences, Guangzhou, China. ²Department of Otolaryngology-Head and Neck Surgery, Guangdong Provincial People's Hospital (Guangdong Academy of Medical Sciences), Southern Medical

University, Guangzhou 510000, Guangdong, China. ³Department of Oncology, Zhujiang Hospital, Southern Medical University, Guangzhou, China.

Received: 12 September 2023 Accepted: 1 July 2024

Published online: 14 October 2024

References

- Breiteneder H, Peng Y-Q, Agache I, Diamant Z, Eiwegger T, Fokkens WJ, et al. Biomarkers for diagnosis and prediction of therapy responses in allergic diseases and asthma. *Allergy*. 2020;75:3039–68. <https://doi.org/10.1111/all.14582>.
- Samitas K, Carter A, Kariyawasam HH, Xanthou G. Upper and lower airway remodelling mechanisms in asthma, allergic rhinitis and chronic rhinosinusitis: the one airway concept revisited. *Allergy*. 2018;73:993–1002. <https://doi.org/10.1111/all.13373>.
- Barbarot S, Auziere S, Gadkari A, Girolomoni G, Puig L, Simpson EL, et al. Epidemiology of atopic dermatitis in adults: results from an international survey. *Allergy*. 2018;73:1284–93. <https://doi.org/10.1111/all.13401>.
- Fokkens WJ, Lund VJ, Hopkins C, Hellings PW, Kern R, Reitsma S, et al. European position paper on rhinosinusitis and nasal polyps 2020. *Rhinology*. 2020;58:1–464. <https://doi.org/10.4193/Rhin20.600>.
- Zheng T, Yu J, Oh MH, Zhu Z. The atopic march: progression from atopic dermatitis to allergic rhinitis and asthma. *Allergy Asthma Immunol Res*. 2011;3:67–73. <https://doi.org/10.4168/air.2011.3.2.67>.
- Murrison LB, Brandt EB, Myers JB, Hershey GKK. Environmental exposures and mechanisms in allergy and asthma development. *J Clin Invest*. 2019;129:1504–15. <https://doi.org/10.1172/JCI124612>.
- Raimondo A, Lembo S. Atopic dermatitis: epidemiology and clinical phenotypes. *Dermatol Pract Concept*. 2021;11:e2021146. <https://doi.org/10.5826/dpc.1104a146>.
- Sancllemente G, Hernandez N, Chaparro D, Tamayo L, Lopez A, Colombian Atopic Dermatitis Research Group. Epidemiologic features and burden of atopic dermatitis in adolescent and adult patients: a cross-sectional multicenter study. *World Allergy Organ J*. 2021;14:100611. <https://doi.org/10.1016/j.waojou.2021.100611>.
- Li Y-T, Hou M-H, Lu Y-X, Chen P-R, Dai Z-Y, Yang L-F, et al. Multimorbidity of allergic conditions in urban citizens of southern china: a real-world cross-sectional study. *J Clin Med*. 2023;12:2226. <https://doi.org/10.3390/jcm12062226>.
- Shirai Y, Nakanishi Y, Suzuki A, Konaka H, Nishikawa R, Sonehara K, et al. Multi-trait and cross-population genome-wide association studies across autoimmune and allergic diseases identify shared and distinct genetic component. *Ann Rheum Dis*. 2022;81:1301–12. <https://doi.org/10.1136/annrheumdis-2022-222460>.
- Sordillo JE, Zhou Y, McGeachie MJ, Ziniti J, Lange N, Laranjo N, et al. Factors influencing the infant gut microbiome at age 3–6 months: findings from the ethnically diverse Vitamin D Antenatal Asthma Reduction Trial (VDAART). *J Allergy Clin Immunol*. 2017;139:482–491.e14. <https://doi.org/10.1016/j.jaci.2016.08.045>.
- Barshad G, Webb LM, Ting H-A, Oyesola OO, Onyekwere OG, Lewis JJ, et al. E-protein inhibition in ILC2 development shapes the function of mature ILC2s during allergic airway inflammation. *J Immunol*. 1950;2022(208):1007–20. <https://doi.org/10.4049/jimmunol.2100414>.
- Kong WS, Tsuyama N, Inoue H, Guo Y, Mokuda S, Nobukiyo A, et al. Long-chain saturated fatty acids in breast milk are associated with the pathogenesis of atopic dermatitis via induction of inflammatory ILC3s. *Clin Rep*. 2021;11:13109. <https://doi.org/10.1038/s41598-021-92282-0>.
- Rochman M, Kartashov AV, Caldwell JM, Collins MH, Stucke EM, Kc K, et al. Neurotrophic tyrosine kinase receptor 1 is a direct transcriptional and epigenetic target of IL-13 involved in allergic inflammation. *Mucosal Immunol*. 2015;8:785–98. <https://doi.org/10.1038/mi.2014.109>.
- Guttman-Yassky E, Bissonnette R, Ungar B, Suárez-Fariñas M, Ardeleanu M, Esaki H, et al. Dupilumab progressively improves systemic and cutaneous abnormalities in patients with atopic dermatitis. *J Allergy Clin Immunol*. 2019;143:155–72. <https://doi.org/10.1016/j.jaci.2018.08.022>.
- Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal*. 2013;6:pl1. <https://doi.org/10.1126/scisignal.2004088>.

17. Tang Z, Kang B, Li C, Chen T, Zhang Z. GEPIA2: an enhanced web server for large-scale expression profiling and interactive analysis. *Nucleic Acids Res.* 2019;47:W556–60. <https://doi.org/10.1093/nar/gkz430>.
18. Tang G, Cho M, Wang X. OncoDB: an interactive online database for analysis of gene expression and viral infection in cancer. *Nucleic Acids Res.* 2022;50:D1334–9. <https://doi.org/10.1093/nar/gkab970>.
19. Wang S, Xiong Y, Zhao L, Gu K, Li Y, Zhao F, et al. UCSCXenaShiny: an R/CRAN package for interactive analysis of UCSC Xena data. *Bioinformatics.* 2022;38:527–9. <https://doi.org/10.1093/bioinformatics/btab561>.
20. Clough E, Barrett T. The gene expression omnibus database. *Methods Mol Biol.* 2016;1418:93–110. https://doi.org/10.1007/978-1-4939-3578-9_5.
21. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-seq and microarray studies. *Nucleic Acids Res.* 2015;43: e47. <https://doi.org/10.1093/nar/gkv007>.
22. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics.* 2010;26:139–40. <https://doi.org/10.1093/bioinformatics/btp616>.
23. McCarthy DJ, Chen Y, Smyth GK. Differential expression analysis of multi-factor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Res.* 2012;40:4288–97. <https://doi.org/10.1093/nar/gks042>.
24. Chen Y, Lun ATL, Smyth GK. From reads to genes to pathways: differential expression analysis of RNA-Seq experiments using Rsubread and the edgeR quasi-likelihood pipeline. *F1000Research.* 2016;5:1438. <https://doi.org/10.12688/f1000research.8987.2>.
25. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA.* 2005;102:15545–50. <https://doi.org/10.1073/pnas.0506580102>.
26. Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, et al. clusterProfiler 4.0: a universal enrichment tool for interpreting omics data. *Innovation.* 2021;2:100141. <https://doi.org/10.1016/j.xinn.2021.100141>.
27. Hänzelmann S, Castelo R, Guinney J. GSEA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinf.* 2013;14:7. <https://doi.org/10.1186/1471-2105-14-7>.
28. Sturm G, Finotello F, List M. Immunedconv: an R package for unified access to computational methods for estimating immune cell fractions from bulk RNA-sequencing data. *Methods Mol Biol.* 2020;2120:223–32. https://doi.org/10.1007/978-1-0716-0327-7_16.
29. Sturm G, Finotello F, Petitprez F, Zhang JD, Baumbach J, Fridman WH, et al. Comprehensive evaluation of transcriptome-based cell-type quantification methods for immuno-oncology. *Bioinformatics.* 2019;35:i436–45. <https://doi.org/10.1093/bioinformatics/btz363>.
30. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, et al. Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods.* 2015;12:453–7. <https://doi.org/10.1038/nmeth.3337>.
31. Becht E, Giraldo NA, Lacroix L, Buttard B, Elarouci N, Petitprez F, et al. Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. *Genome Biol.* 2016;17:218. <https://doi.org/10.1186/s13059-016-1070-5>.
32. Petitprez F, Levy S, Sun C-M, Meylan M, Linhard C, Becht E, et al. The murine Microenvironment Cell Population counter method to estimate abundance of tissue-infiltrating immune and stromal cell populations in murine samples using gene expression. *Genome Med.* 2020;12:86. <https://doi.org/10.1186/s13073-020-00783-w>.
33. Finotello F, Mayer C, Plattner C, Laschober G, Rieder D, Hackl H, et al. Molecular and pharmacological modulators of the tumor immune contexture revealed by deconvolution of RNA-seq data. *Genome Med.* 2019;11:34. <https://doi.org/10.1186/s13073-019-0638-6>.
34. Aran D, Hu Z, Butte AJ. xCell: digitally portraying the tissue cellular heterogeneity landscape. *Genome Biol.* 2017;18:220. <https://doi.org/10.1186/s13059-017-1349-1>.
35. Racle J, de Jonge K, Baumgaertner P, Speiser DE, Gfeller D. Simultaneous enumeration of cancer and immune cell types from bulk tumor gene expression data. *Elife.* 2017;6: e26476. <https://doi.org/10.7554/eLife.26476>.
36. Hao Y, Hao S, Andersen-Nissen E, Mauck WM, Zheng S, Butler A, et al. Integrated analysis of multimodal single-cell data. *Cell.* 2021;184:3573–3587. e29. <https://doi.org/10.1016/j.cell.2021.04.048>.
37. Wickham H. Modelling for visualisation. In: Wickham H, editor. *Ggplot2 elegant graphics for data analysis*. Cham: Springer International Publishing; 2016. p. 221–40. https://doi.org/10.1007/978-3-319-24277-4_11.
38. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics.* 2016;32:2847–9. <https://doi.org/10.1093/bioinformatics/btw313>.
39. Gu Z. Complex heatmap visualization. *iMeta.* 2022;1: e43. <https://doi.org/10.1002/imt2.43>.
40. Yu G, Hu E, Gao C-H. enrichplot: visualization of functional enrichment result. n.d.
41. Wei T, Yu G. GseaVis: implement for “GSEA” enrichment visualization. 2022.
42. Wei T, Simko V, Levy M, Xie Y, Jin Y, Zemla J, et al. corrrplot: visualization of a correlation matrix. 2021.
43. Chang W, Cheng J, Allaire JJ, Sievert C, Schloerke B, Xie Y, et al. shiny: web application framework for R. 2024.
44. Xie Y, Cheng J, Tan X, Allaire JJ, Gillich M, Ellis GF, et al. DT: A wrapper of the JavaScript library “DataTables”. 2024.
45. Brożek JL, Bousquet J, Agache I, Agarwal A, Bachert C, Bosnic-Anticevich S, et al. Allergic rhinitis and its impact on asthma (ARIA) guidelines-2016 revision. *J Allergy Clin Immunol.* 2017;140:950–8. <https://doi.org/10.1016/j.jaci.2017.03.050>.
46. Furusawa H, Cardwell JH, Okamoto T, Walts AD, Konigsberg IR, Kurche JS, et al. Chronic hypersensitivity pneumonitis, an interstitial lung disease with distinct molecular signatures. *Am J Respir Crit Care Med.* 2020;202:1430–44. <https://doi.org/10.1164/rccm.202001-01340C>.
47. Kim H-M, Lee CH, Rhee C-S. Histamine regulates mucin expression through H1 receptor in airway epithelial cells. *Acta Otolaryngol.* 2012;132(Suppl 1):S37–43. <https://doi.org/10.3109/00016489.2012.661075>.
48. Parrish A, Boudaud M, Kuehn A, Ollert M, Desai MS. Intestinal mucus barrier: a missing piece of the puzzle in food allergy. *Trends Mol Med.* 2022;28:36–50. <https://doi.org/10.1016/j.molmed.2021.10.004>.
49. Schwalm K, Stevens JF, Jiang Z, Schuyler MR, Schrader R, Randell SH, et al. Expression of the proapoptotic protein Bax is reduced in bronchial mucous cells of asthmatic subjects. *Am J Physiol Lung Cell Mol Physiol.* 2008;294:L1102–1109. <https://doi.org/10.1152/ajplung.00424.2007>.
50. Tomazic PV, Darnhofer B, Birner-Gruenberger R. Nasal mucus proteome and its involvement in allergic rhinitis. *Expert Rev Proteomics.* 2020;17:191–9. <https://doi.org/10.1080/14789450.2020.1748502>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.