Characterization of the Immune Microenvironment and Identification of Biomarkers in Chronic Rhinosinusitis with Nasal Polyps Using Single-Cell RNA Sequencing and Transcriptome Analysis

Yakun Wang1,*, Xinyu Song2,*, Mulan Jin1, Jun Lu1

1Department of Pathology, Beijing Chaoyang Hospital, Capital Medical University, Beijing, 100020, People’s Republic of China; 2Department of Otorhinolaryngology, Head and Neck Surgery, Beijing Chaoyang Hospital, Capital Medical University, Beijing, 100020, People’s Republic of China

*These authors contributed equally to this work

Correspondence: Mulan Jin; Jun Lu, Department of Pathology, Beijing Chaoyang Hospital, Capital Medical University, Beijing, 100020, People’s Republic of China, Email jinmulan2022@126.com; Lujun11101@126.com

Purpose: Chronic rhinosinusitis is a prevalent condition in the field of otorhinolaryngology; however, its pathogenesis remains to be elucidated. The immunological defense of the nasal mucosa is significantly influenced by dendritic cells (DCs). We identified specific biological indicators linked to DCs and explored their significance in cases of chronic rhinosinusitis with nasal polyps (CRSwNP).

Patients and Methods: We categorized cells using single-cell RNA (scRNA) sequencing, and combined transcriptome sequencing was used to identify potential candidate genes for CRSwNP. We selected three biomarkers based on two algorithms and performed enrichment and immune correlation analyses. Biomarkers were verified using training and validation sets, receiver operating characteristic curves, immunohistochemistry, and quantitative real-time reverse-transcription PCR (qRT-PCR). Variations in biomarker expression were validated using pseudotime analysis. The networks of competing transcription factor (TF)-mRNA and competing endogenous RNA (ceRNA) were established, and the protein drugs associated with these biomarkers were predicted.

Results: Both scRNA-seq and transcriptome data showed that DCs immune infiltration was higher in the CRSwNP group than in the control group. Three DC-related biomarkers (NR4A1, CLEC4G, and CD163) were identified. In CRSwNP, NR4A1 expression decreased, whereas CLEC4G and CD163 expression increased. All biomarkers were shown to be involved in immunological and metabolic pathways by enrichment analysis. These biomarkers were associated with γδ T cells, effector memory CD4+ T cells, regulatory T cells, and immature DCs. According to pseudotime analysis, NR4A1 and CD163 expression decreased from high to low, whereas CLEC4G expression remained low.

Conclusion: We screened and identified potential DC-associated biomarkers of CRSwNP progression by integrating scRNA-seq with whole transcriptome sequencing. We analyzed the biological pathways in which they were involved, explored their molecular regulatory mechanisms and related drugs, and constructed ceRNA, TF-mRNA, and biomarker–drug networks to identify new CRSwNP treatment targets, laying the groundwork for the clinical management of CRSwNP.

Keywords: CRSwNP, dendritic cell, single-cell sequencing, transcriptomic sequencing, biomarker

Introduction

Chronic rhinosinusitis (CRS) is a common disease in the field of otorhinolaryngology, however, information about its pathogenesis remains to be elucidated. Different pathogenic mechanisms of CRS mediate different types of immunopathologies (“disease endotypes”), which produce different clinical manifestations known as disease phenotypes. Clinically, CRS is frequently classified into two distinct phenotypes based on the presence or absence of nasal polyps: CRS sans nasal polyps and CRS with nasal polyps (CRSwPNPs). The latter is divided into eosinophilic CRSwNP (eCRSwNP) and non-eosinophilic CRSwNP (neCRSwNP). The inflammatory microenvironment in CRSwNP is
believed to contribute to the disease among other factors including inflammation and allergies.\(^3\) Allergic rhinitis is a chronic inflammatory disease of the nasal mucosa. The long-term existence of allergic rhinitis, so that the inflammatory response delayed, which can lead to chronic sinusitis and nasal polyp formation.\(^4\) Asthma is the most important and common complication of allergic rhinitis. Epidemiological studies and clinical trials have found a correlation between the three diseases. Allergic rhinitis, sinusitis and asthma are three different manifestations of a common pathological process with similar pathological mechanisms. As a result of being affected by allergies, eosinophils are thought to damage the epithelium by releasing cytokines and other pro-inflammatory proteins. On this side, the damaged cytokines and chemokines, further attract eosinophils, play a central role in this process. Chronic airway mucosal inflammation and remodelling play a key integrative role in these diseases.\(^5\) Increasing evidence suggests that CRSwNP does not occur as a solitary disease, and instead, it presents as a heterogeneous condition with different clinical phenotypes and treatment prognoses driven by different pathogenic mechanisms.\(^6\)

Although the exact cause of the persistent inflammation associated with CRSwNP is unknown, it is possible that it arises from a multitude of interactions involving environmental factors and the host immune system.\(^7\) Infection, metaplasia, mucosal cilia damage, and anatomical variations of the nose-sinus are usually considered to be causative factors. Bacteria can directly trigger inflammation and exacerbate the pathological process of CRS. \(S. \) \textit{aureus} is thought to play an important role in the immunopathogenesis of CRSwNP and secretes a variety of exotoxins. It acts as a superantigen and produces toxins that regulate the immune system. These exotoxins enhance the invasive capacity and persistence of the bacteria by hindering the body’s immune system response against them. Among which interleukins (LukED) may serve as prototypical targets for CRSwNP.\(^8\) Additionally, epigenetics contributes to contribute to a deeper understanding of the pathogenesis of CRSwNP by regulating the expression and activity of host genes, the epigenome can undergo rapid DNA methylation changes when the environment is altered. In addition to DNA methylation, histone modifications and miRNA aberrations can contribute to disease by regulating downstream genes, cytokines, transcription factors, proteins, etc.\(^9\) The immune system is a dynamic and complete network comprising multiple types of immune cells distributed throughout the body in different organs and tissues, which work together to maintain intra-tissue homeostasis and participate in protective immunity.\(^10-12\) Numerous investigations have delved into the makeup of immune cells within human tissues and organs through single-cell transcriptome sequencing (scRNA-seq) and have identified immune cells unique to specific tissues.\(^13,14\) A strong T2-type immune response is a typical feature of CRSwNP, and many studies have identified multiple specific cellular subpopulations and molecules that contribute to the heterogeneity of CRS subtypes underneath the T2-type immune response by scRNA-seq. Among them, ALOX15+ cDC2s and macrophages play a key role in the pathogenesis of type 2 immunity in eCRSwNP.\(^15\) In addition, the diversity of epithelial cells, which are central to the pathogenesis of CRSwNP, decreases with changes in the T2 immune environment. This leads to a severely impaired physiological barrier function. The pathophysiology of CRSwNP is further exacerbated by mast cell (MC) cells as an important source of high levels of highly concentrated pro-inflammatory mediators.\(^16\) In turn, basal cells in CRSwNP develop a “memory” of long-term exposure to an inflammatory T2 environment, which in turn leads to a “memory” of an entire cellular ecosystem that deviates from productive differentiation and propagates disease.\(^17\) ScRNA-seq quantifies gene expression, functional enrichment, and metabolic pathways of mRNAs in single cells, which can evaluate cellular heterogeneity in research fields such as early embryonic development, stem cells, cancer, and immunology and represents a potent apparatus for exploring gene expression at the single cell level.\(^18-20\) Dendritic cells (DCs) are specifically responsible for presenting antigens and constitute the initial front line of primary and innate cellular immune reactions.\(^21\) DCs are derived from continuously proliferating progenitor cells within the bone marrow and precursor cells relocate into the vascular system, where they undergo the process of monocyte transformation.\(^22,23\) Depending on further antigenic stimulation, immature DCs typically enter the regional lymph nodes and spleen, transforming into immunologically mature forms that stimulate T and B lymphocytes for the generation of immune responses tailored to specific antigens.\(^24\) DC-mediated immune responses lead to the simultaneous expression of pro-inflammatory cytokines, namely IL-1, TNF-α, and growth factors (TGF-α, PDGF, EGF, and FGF) in peculiar patterns of massive release.\(^24-26\) However, the pathogenic mechanisms underlying DC accumulation, pro-inflammatory cytokine overexpression, and nasal polyp formation remain unclear. We performed the study according to
Ordovas-Montanes et al\textsuperscript{16} and Zhang et al\textsuperscript{15} annotation results of DC cells from tissue samples of CRSwNP identified cell types and corresponding maker genes.

Here, the objective of this investigation was to identify a population of DCs with dysregulated gene expression in CRSwNP and identify potential biomarkers linked to DCs during the course of CRSwNP development, to construct a DC-centered global regulatory network. By analyzing the biological pathways, molecular regulatory mechanisms, and related drugs of DC-related biomarkers in CRSwNP to uncover vital pathways that could lead to the discovery of candidate genes for drug-targeting thus supporting the clinical management of CRSwNP.

**Materials and Methods**

**Study Design and Population**

**Ethical Statement**

The Beijing Chaoyang Hospital’s Ethics Committee’s Institutional Review Board (No. 2020-3-16-18 and date of approval: 2020-3-19) gave its approval to this work. All experiments involving human participants adhered to the institutional ethical guidelines and followed the principles of the Declaration of Helsinki. Before participating in the study, all participants involved provided written informed permission and finished informed consent procedure.

**Subject Recruitment and Sample Collection**

Between July 2021 and July 2022, the patients who received nasal polyp surgery or surgical correction of a deviated septum were enrolled. The diagnosis of CRSwNP was in accordance with the European Position Paper on Rhinosinusitis and Nasal Polyps 2020.\textsuperscript{27} Participants who had fungal sinusitis, antrochoanal polyps, unilateral polyps, cystic fibrosis or primary ciliary dyskinesia were excluded. And CRSwNP patients who took oral glucocorticoids, antimicrobials, antileukotrienes, and antihistamines 4 weeks prior to sample collection were excluded. Deviated nasal septum patients with coexisting asthma, allergic rhinitis, and/or aspirin sensitivity were also excluded. A total of 72 patients were enrolled, including CRSwNP group (n=36) and control group (n=36). Three samples were randomly selected from each group for scRNA-seq analysis: three mucosal tissue samples from nasal polyps (NP group), three nasal mucosa tissue samples adjacent to nasal polyps (NM group) from the CRSwNP group, and three nasal mucosa tissue samples from the control group (CM group).

To conduct transcriptome sequencing, the CRSwNP model’s training dataset was created using eight nasal mucosal tissue samples from the nasal polyp group and eight nasal mucosal tissue samples from the control group. Furthermore, data from GSE136825, encompassing 42 polyp samples from CRSwNP patients and 28 turbinate samples from non-CRS controls, were retrieved from the Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/) and served as the validation dataset.\textsuperscript{28} Tissue samples from 15 patients with NP in the CRSwNP group and 15 nasal mucosal tissue samples from the control group were processed for immunohistochemistry (IHC) analysis to verify the analytical results. Tissue samples from ten patients with NP in the CRSwNP group and ten nasal mucosal tissue samples from the control group were used in quantitative real-time reverse transcription polymerase chain reaction (qRT-PCR) analysis to validate the results.

**Single-Cell Sequencing Analysis**

**Single-Cell Sequencing**

The concentration and quality of the single-cell suspensions were strictly controlled based on the source and characteristics of the tissue or cell samples. The obtained tissue samples were washed 3 times in DPBS and transferred to centrifuge tube with 10 cm surgical scissors cut into small pieces. The samples were incubated at 37°C in a constant temperature water bath with the enzymatic digestion kit 1 hour. The size, cell number and viability of the tissue block were observed every 30 min during dissociation. Generally, the number of cells reached 10\textsuperscript{6}, the digestion solution became cloudy, and the incubation was terminated when the tissue block disappeared. The cells were filtered through a 40 µm sieve. The cell suspension was centrifuged at 300rcf for 5 min at 4°C, and the cells were resuspended. The cells were incubated at 4°C for 5 min, and immediately after the end, the cells were centrifuged at 300rcf for 5min at 4°C. The cells were fully resuspended by adding 1mL of resuspension. The target concentration was 700–1200 cells /µL. The cell
activity was more than 90% and the agglomeration was less than 15%. Once the final cell concentration and activity were reached, the cells were placed on ice and the on-machine experiments of the 10x Genomics single-cell transcriptome chip were performed within 30 min. Cell quality inspection standard: cell concentration: 700–1200 cells/microliter. Cell volume:>60 microliters. Cell activity:> 90%. Agglomeration rate:< 15%. Cell diameter: <30 microns. It is important to note that because of the difficulty in isolating granulocytes from frozen samples, fresh tissue samples were selected first and should be processed immediately after granulocyte collection, no more than 2 hours later. Increasing the number of PCR cycles Increasing the number of PCR cycles by two during cDNA amplification of nonenriched samples may help to keep granulocyte-specific transcripts above the background. Unsupervised analysis of samples by Cell Ranger software may filter out those neutrophils that do not have much mRNA. We used the “–force-cells” parameter and set its value to target cell recovery.29,30

Chromium Next GEM Single-Cell 3′ Reagent Kits v3.1 were employed in conjunction with the Chromium Controller from 10×Genomics to generate libraries for scRNA-seq. Cell Ranger software and STRA aligner were used to analyze the sequencing data. Specifics regarding the preparation of sequencing libraries and preprocessing of scRNA-seq data can be found in the Supplementary Material 1. Seurat (4.3.0) software was used to cluster the data and its dimensionality was reduced using principal component analysis (PCA). Uniform manifold approximations and projections were used for visualization and presentation. SingleR (version 2.0.0) was used for automatic annotation of cellular taxa and manual annotation corrections. Cell types were corrected using the Cell Marker (http://bio-bigdata.hrbmu.edu.cn/CellMarker/).

Identifying the Differentially Expressed Genes Related to DCs via Single-Cell Analysis
First, quality control procedures were carried out using the “Seurat” R platform (version 4.3.0) based on scRNA-seq data with 500 < nFeature RNA < 7,000, nCount RNA < 50,000, and percent mt < 20%.31 Data were standardized through the utilization of the Normalize Date function on the “Seurat” R package (version 4.3.0). The top 2,000 high-variance genes were extracted using the FindVariable Features function. Clustering analyses were performed on the cells remaining after data integration and PCA dimensionality reduction. Subsequently, we used Human Primary Cell Atlas Data as a reference dataset and annotated the cell types via the “SingleR” R package (version 2.0.0), with the expression of marker genes obtained from the published literature.32,33 Moreover, cell to cell communication was evaluated via the “CellChat” R package. Myeloid cells were re-clustered and re-annotated. Differences in DCs among the NP, NM, and CM groups were assessed using the χ²-test. Differentially expressed genes (DEGs) were detected between NP and CM, and between NP and NM groups employing FindMarkers within the “Seurat” R package (version 4.3.0) with P<0.05 and |log₂FC| > 0.25. These DEGs were then intersected to obtain common DEGs, which were re-annotated as DEGs1.

Analysis of Transcriptome Sequencing
Whole Transcriptome Sequencing
Using the TRIzol reagent (Invitrogen, USA), RNA was isolated from the samples and purified in accordance with the manufacturer’s instructions. The total RNA’s amount and purity were verified using a NanoDrop ND-1000 (NanoDrop, USA). Additionally, RNA integrity was evaluated using a Bioanalyzer 2100 (Agilent, USA) with a focus on achieving RIN values exceeding 7.0 and validated by denaturing agarose gel electrophoresis. Illumina NovaSeq 6000 (LC Bio-Technology CO., Ltd. Hangzhou, China) was used for bipartite sequencing according to standard procedures. The sequencing mode used was PE150. Reads with adapter contamination were removed using the Fastp program (https://github.com/OpenGene/fastp). Sequencing data were compared to GRCh38_genome_v33 using HISAT2 (https://ccb.jhu.edu/software/htsat2) to obtain the read alignment file (in.bam file format). With the default settings, StringTie (https://ccb.jhu.edu/software/stringtie) was used to generate the mapped readings from each sample. Transcripts from all samples were amalgamated to reconstruct an all-encompassing transcriptome using gffcompare (https://github.com/gpertea/gffcompare). Genes and transcripts were assembled and quantified as total_exon_fragments/mapped_reads (in millions) × exon_length (in kilobases) using the StringTie software (https://ccb.jhu.edu/software/htsat2),(DEGs) between samples were scrutinized using the R package “DESeq2” (version 1.34.0), with criteria including P<0.05 and |log2FC|>0.5.34 Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses were conducted for genes through the DAVID software (https://david.ncifcrf.gov/).
Analysis of Immune Cell Infiltration
The infiltrating abundance of 28 immune cells in the CRSwNP and control groups was evaluated using a single-sample gene set enrichment analysis. Discrepancies between the CRSwNP and control groups were compared using the Wilcoxon test. Moreover, the correlation between immune cells was evaluated by Spearman correlation analysis in the “corrplot” R package (version 0.92).^{35}

Creation of a Weighted Gene Co-Expression System
To uncover modules linked to DCs, co-expression of weighted genes was executed based on transcriptome sequencing data. Outliers were cleared using sample clustering. To ensure that the gene interactions conformed as much as possible to a scale-free distribution, we performed soft threshold (β) determination. Next, the similarity among genes was calculated according to adjacency, and a gene dendrogram including different modules was created using the degree of dissimilarity. The relationships between the modules and DCs were calculated using Pearson’s correlation analysis.^{36} Modules with an absolute correlation value greater than 0.5 were most positively or negatively associated with DCs and were therefore selected for further study; the genes in these modules were termed DC-related genes (DCRGs).

Identification of Candidate Genes
First, the DEGs between the CRSwNP and control groups (DEGs2) were identified using the “DESeq2” R package (version 1.34.0) ($P<0.05$) and $|\log_2FC|>0.5$.^{37} DEGs1, DCRGs, and DEGs2 were then intersected to obtain intersecting genes via the “ggvenn” R package (version 0.1.9), which were recorded as candidate genes. The biological functional enrichment analysis of candidate genes based on the GO and KEGG databases was carried out using the “clusterProfiler” R package (version 4.7.1) and “org.Hs.eg.db” R package (version 3.16.0) ($P<0.05$).^{38}

Selection of Biomarkers
Support vector machine-recursive feature elimination (SVM–RFE) was implemented using the “caret” R platform and XGBoost (through the “xgboost” R package) were employed to identify characteristic genes.^{39} We then established an intersection between the two machine-learning-obtained characteristic genes to identify common characteristic genes.

Expression Analysis of Biomarkers
Analysis of Tissue Expression and Validation of External Data Sets
The expression trends of the common characteristic genes in the training and validation sets were examined. The diagnostic potential of these shared characteristic genes for CRSwNP was evaluated by constructing receiver operating characteristic curves through the “pROC” R package (version 1.18.0).^{40} In both the training and validation datasets, genes exhibiting statistically significant expression differences and displaying consistent expression patterns between the CRSwNP and control groups were considered. Additionally, the area under the curve (AUC) values higher than 0.7 were considered biomarkers.

Single-Cell Analysis of Biomarkers and Pseudotime Analysis
To visualize the expression levels of the three hub genes (NR4A1, CLEC4G, and CD163) across various cell types, the DotPlot function from the “Seurat” R package was employed. To conduct a more comprehensive analysis of biomarker expression in different tissues, the expression of these biomarkers was examined using the Human Protein Atlas (https://www.proteinatlas.org/). Additionally, the expression of biomarkers in different cell types was assessed via the DotPlot function in the “Seurat” R package. We further selected cells in which all biomarkers were expressed for pseudotime analysis using the “monocle3” R package. Finally, the correlation coefficients calculated by Spearman’s technique were used to evaluate the relationship between biomarkers and immune cells.

Immunohistochemical Analysis
Tissue samples underwent a series of processing steps, including dehydration, paraffin embedding, rotary microtome sectioning, and placement onto glass slides. The sections were dried for approximately 1 hour at 65°C. The tissue
sections were deparaffinized using xylene and ethanol, immersed in a citrate solution, and heated for 20 minutes in a microwave oven. To prevent non-specific antibody binding, bovine serum albumin was used as a blocking agent. Subsequently, the tissue sections were incubated with primary antibodies (anti-NR4A1, 1:200 dilution; Proteintech; anti-CLEC4G, 1:100 dilution; anti-CD163, 1:1,000 dilution; Abcam, Cambridge, MA USA) in a humidified chamber at 4°C overnight. Subsequently, the cells were incubated with the secondary antibodies for 40 minutes at room temperature. The ImageJ software was used to analyze IHC staining for positivity (Bethesda, MD, USA).

qRT-PCR
RNA from tissues was purified using TRIzol (Invitrogen, USA) and was reverse transcribed using a reverse transcription kit (Transgene, Beijing, China). qRT-PCR analyses were performed using SYBR Green (Transgene, Beijing, China) on an Applied Biosystems 7500 Real-Time PCR System (Life Technologies, Gaithersburg, MD, USA), and gene expression levels were normalized to endogenous GAPDH. The primer sequences (5'-3') used to amplify NR4A1 were forward, GCTTCGGGGACTGGATTGA and reverse, GGTGATGAGGACAAGGGCAGA. The primers used to amplify CLEC4G were forward, CGACCTGCTGAGGACAAACG and reverse, AGTCTCCGACCTCCTCCTTCA. The primers used to amplify CD163 were forward, CCAAAATCCAGGCAACAAACA and reverse, CTGTGGGCTGAGCAGGTAATT. The primers used to amplify GAPDH were forward, GGAAGCTTGTCATCAATGGAAATC and reverse TGATGACCCTTTTGGCTCCC.

Enrichment Analysis of Gene Sets
To perform an in-depth investigation into the biological functions and signaling pathways associated with the biomarkers, the median value of each biomarker’s expression was used as a criterion to divide the training dataset samples into groups with high and low expression, and subsequent differential analysis was conducted. Genes were ranked by logFC values, and gene set enrichment analysis was performed, considering adj \( P<0.05 \) and (|NES|) \( \geq 1 \) as criteria.

Correlation Analysis Between Biomarkers and Immune Cells
Spearman correlation between gene expression and the fraction of immune cells was performed using the R package ggplot2 to illustrate the relationship between immune cell infiltration and the levels of biomarker expression. Initially, data consolidation and normalization were performed by processing the gene expression matrix and the list of results related to immune cell infiltration. Subsequently, correlation tests were performed for various immune cell types within the dataset, and the outcomes were visualized using correlation scatterplots. Finally, the relationship between the biomarkers and immune cells was depicted using a lollipop plot.

Creation of Regulatory Networks
To further investigate the mechanisms underlying biomarkers in CRSwNP, we created competing endogenous RNA (ceRNA) and TF-mRNA networks. First, the miRNAs regulating biomarkers were predicted using the MicroCosm (http://www.ebi.ac.uk/enright-srv/microcosm/htdocs/targets/v5/) and miRDB (http://www.mirdb.org/) databases. The miRNAs sourced from the two databases were compared to identify shared miRNAs, which were used to construct a ceRNA network. These common miRNAs were further employed in the starBase database (https://starbase.sysu.edu.cn/) to predict lncRNAs involved in regulatory interactions. Moreover, we selected lncRNAs with clipExpNum > 4 to construct the ceRNA network. TFs were predicted using the NetworkAnalyst database (https://www.networkanalyst.ca/). Finally, Cytoscape was used to visualize the regulatory networks.

Drug Prediction and Disease Prediction
The Drug-Gene Interaction Database (https://dgidb.genome.wustl) and the Comparative Toxicogenomics Database (CTD) (http://ctd.mdibl.org) were used to predict the correlation between drugs and diseases (ear, nose, and throat diseases) and biomarkers. The Cytoscape software was used to create a visual representation of the biomarker-drug network.
Statistical Analysis
Quantitative variables were expressed as median value and interquartile range and the Mann Whitney U-test was utilized. Categorical variables were presented as value and percentage with the chi-square test used. The significant level of differences between groups was ascertained by using the student’s t test to assess the expression levels of biomarkers. Based on the empirical Bayes method, the R package “DESeq2” was used for the analysis of common transcriptome differential genes. Spearman correlation analysis was used to explore the relationship between two variables, such as the correlation between immune cells. The statistical software SPSS (Version 21, Chicago, IL, USA), GraphPad Prism (Version 9.3.0, GraphPad Software Inc., San Diego, USA), or R software (Version 4.2.3) were used when appropriate. Statistical difference was defined as a two-sided P value<0.05. (ns, no significant, * P < 0.05, ** P < 0.01, *** P < 0.001, **** P < 0.0001).

Results
Patient Characteristics
The demographic characteristics were described in Table 1. According to the inclusion and exclusion criteria, a total of 72 patients were enrolled, including CRSwNP group (n=36) and control group (n=36). The median age was 44.50 and 49.00, respectively. There were 21 males and 15 females in the CRSwNP group, while 16 males and 20 females in the control group. Based on the exclusion criteria, no asthma, allergic rhinitis, or NERD were in the control group. In contrast, there were 5 cases with asthma (13.9%), 19 with allergic rhinitis (52.8%), and 1 with NERD (2.8%) in the CRSwNP group. A significant difference was found in allergic rhinitis between the two groups (P<0.001).

Identification of Common DEGs Between Different Groups
After quality control, 53,571 cells and 29,931 genes were used for subsequent analyses. After data normalization, we characterized the top 2,000 genes with elevated variance and the top ten genes. Once the data were integrated across samples, PCA downscaling analysis was performed, and 20 principal components were selected for subsequent analyses. The grouping of the two principal components was relatively high in the CM, NM, and NP groups. A total of 21 cell clusters were identified; however, cell cluster 9 was present only in the NP and NM groups and cell cluster 20 was present only in the NP and CM groups. A total of 21 clusters of cells were identified as epithelial cells, endothelial cells, fibroblasts, smooth muscle cells, B cells, T cells, and myeloid cells. Cluster 9 was epithelial cells; cluster 20 was fibroblasts. The number of epithelial cells was highest level in the NM and CM groups, whereas the percentage of endothelial cells was highest in the NP group. Cell-cell communication analysis demonstrated that all seven cell types interacted with each other; however, the strength of the interaction varied from cell to cell. We subsequently characterized the clusters of myeloid cells. Eleven cell clusters were identified and annotated into four cell types: mast cells, neutrophils, DCs, and macrophages. Differential analysis showed that the distribution of DCs differed considerably among the NP, NM, and CM groups (P=1.48×10^-21). Therefore, we screened the DEGs for DCs in NP vs NM and NP vs CM. In the NP vs

| Table 1 Overall Patient Demographic Characteristics and Associated Clinical Features |
|---------------------------------|----------------|----------------|----------------|
| Characteristics                  | CRSwNP (n=36) | Control (n=36) | P Value        |
| Age, median (IQR)               | 44.50 (38.00,52.75) | 49.00 (39.00,58.75) | 0.224          |
| Sex (M/F)                       | 21/15          | 16/20          | 0.238          |
| Asthma                          | 5 (13.9%)      | 0              | 0.054          |
| Allergic rhinitis               | 19 (52.8%)     | 0              | <0.001***      |
| NERD                            | 1 (2.8%)       | 0              | 1.000          |

Note: ***p < 0.001.
Abbreviations: CRSwNP, chronic rhinosinusitis with nasal polyps; IQR, interquartile range; M, male; F, female; NERD, Non-steroidal anti-inflammatory drugs-exacerbated respiratory disease.
NM group, 779 DEGs were identified, including 548 upregulated and 231 downregulated genes (Figure 2D), whereas 1,339 DEGs were found in the NP vs CM group, including 675 upregulated and 664 downregulated genes (Figure 2D). Following this intersection, we obtained 358 DEGs that could play a key role in CRSwNP progression (Figure 2E).

**Immune Infiltration Analysis Between the CRSwNP and Control Groups**

The heatmap in Figure 3A illustrates the abundance of 28 immune cell types infiltrating both the CRSwNP and control groups. Five immune cell types exhibited statistically significant differences in their infiltration levels between the two groups: immature DCs, γδ T cells, neutrophils, effector memory CD4 T cells, and regulatory T cells (Figure 3B). Moreover, the abundance of these five immune cell types was notably elevated in the CRSwNP group. Immune
Figure 2 Characterization of cell clusters for myeloid cells. (A) UMAP downscaling and cellular taxa. (B) Bubble plots of the expression profiles of diverse marker genes in cellular taxa. (C) Annotation of marker gene cell types. (D) Volcanogram for differential gene analysis. The horizontal coordinate represents the multiplicity of differences in gene expression; the vertical coordinate represents the p-value, which was taken as -log10. In NP vs NM, 779 DEGs were detected, encompassing 548 upregulated and 231 downregulated genes. In NP vs CM, 1,339 DEGs were identified, namely, 675 upregulated and 664 downregulated genes. (E) Screening for shared upregulated and downregulated differential genes. A total of 358 DEGs1 were identified after the intersection.

Journal of Inflammation Research 2024:17
https://doi.org/10.2147/JIR.S440409

DovePress

Wang et al
Figure 3 Analysis of immune cell infiltration. (A) Heat map showing the extent of immune cell infiltration. (B) Distinctions in immune cell abundance between the CRSwNP and control groups. ns, no significant, *P < 0.05. “a” indicates abscissa. The red abscissa indicates that the cell abundance of CRSwNP is up compared to the control group, the gray abscissa indicates no significant.
infiltration of DCs showed the same trend for transcriptome and single-cell sequencing data, both of which were notably elevated in the CRSwNP group compared to the control. The correlations among immune cells are shown in Figure 4. Immature DCs were positively associated with effector memory CD4 T cells (cor = 0.77, \(P = 0.0137\)), suggesting that these two immune cells exhibit synergy in their functions.

### Identification of 18 Candidate Genes

Sample clustering analysis revealed the absence of outlier samples in the training set (Supplementary Figure 3A). When \(\beta = 20\), mean connectivity converged to zero (Supplementary Figure 3B). In total, seven modules with unique colors were characterized: the MEblack module was most positively linked to DCs (\(r = 0.63, P = 0.009\)), whereas the MEblue module was most negatively linked to DCs (\(r = -0.58, P = 0.02\)). The absolute value of the correlation is greater than 0.4 and significant (Figures 5A and B). A total of 1,970 DCRGs were obtained for subsequent analyses. In total, 3,298 DEGs2, including 2,683 upregulated and 615 downregulated ones, were uncovered across the CRSwNP group and its control (Figures 5C and D). Subsequently, 18 candidate genes were identified by intersecting DEGs1, DCRGs, and DEGs2.
Figure 5 Identification of 18 candidate genes. (A) Gene-module tree plot. Each branch symbolizes an individual gene, and the colors below represent distinct co-expression modules. (B) Module-trait relationship chart. Seven modules were formed: MEblack ($r=0.63$, $P=0.009$) and MEblue ($r=-0.58$, $P=0.02$) modules were most significantly related to DCs. (C and D) Volcano and heat maps were used to analyze the differentially expressed genes. (E) Screening for candidate genes.
The GO results were enriched in 186 GO items (including 130 GO BP, 28 GO CC, and 28 GO MF) (Supplementary Table 1), including “cellular response to glucocorticoid stimulus”, “cytosolic small ribosomal subunit”, and “eme binding” (Figures 6A, and B). In total, 30 KEGG pathways exhibited enrichment (Supplementary Table 2), among them, four pathways were notably enriched and activated. These pathways include “coronavirus disease-COVID-19”, “tryptophan metabolism”, “ribosome”, and “collecting duct acid secretion” (Figure 7A). Combined with Z-score >0 (Red) indicates these four pathways are related to “metabolism”, “genetic information processing”, “biological systems”, and “human disorders” (Figure 7B).

Figure 6 (A–B) GO: BP-enriched directed acyclic graphs and chordal graphs. Each box (or oval) displays the GO number and enrichment significance value. Different colors represent different enrichment significance. White indicates no answer, there are missing values or incomplete data in the sample. Red indicates significance.
Identification of Three Biomarkers

The SVM–RFE algorithm was used to identify nine characteristic genes: NR4A1, CLEC4G, CD163, RSRP1, BACH1, F13A1, ITSN1, APBB1IP, and ATP6V1A (Figure 8A). Four characteristic genes were extracted using the XGBoost algorithm: NR4A1, CLEC4G, ZFP36L2, and CD163 (Figure 8B). Three common characteristic genes (NR4A1, CLEC4G, and CD163) were identified using the two machine-learning algorithms (Figure 8C). Within both the training and validation sets, between the CRSwNP and control groups, there were notable differences in the expression levels of all common characteristic genes, with low NR4A1 expression and high CLEC4G and CD163 expression in the CRSwNP group (Figures 8D, and 8E). The AUC levels for all common characteristic genes exceeded 0.7 in both the training and validation sets (Figures 8F, and 8G). In conclusion, we identified three common characteristic genes as potential biomarkers for CRSwNP.

Biomarker Expression Levels

According to the HPA database, all biomarkers were expressed in the liver, lymph nodes, adipose tissue, and urinary bladder (Supplementary Figure 4A–C). All biomarkers were expressed in myeloid cells for all cell types (Figures 9A, and B). Therefore, we selected myeloid cells for the pseudotime analysis. Based on the expression levels of CD86 (a myeloid identifier), the starting stage of the pseudotime period was undifferentiated monocytes in the polyp tissue, which then differentiated into DCs and macrophages (Figures 9C, and D). Moreover, during the pseudotime, the expression of NR4A1 and CD163 showed a trend from high to low, whereas that of CLEC4G remained low (Figure 9E).

IHC analysis was used to measure NR4A1, CLEC4G, and CD163 levels in 15 patients with CRSwNP and 15 control subjects. The IHC staining results indicated that NR4A1 was primarily localized in the cytoplasm. Quantitative analysis
revealed that the proportion of NR4A1-positive cells was significantly higher in the control group than in the CRSwNP group ($P < 0.05$; Figure 10A). Similarly, IHC staining demonstrated that CLEC4G was predominantly expressed in the cytoplasm, with a significantly higher percentage of CLEC4G-positive cells in the CRSwNP group than in the control group ($P < 0.0001$; Figure 10B). Finally, CD163 expression was detected in both the cytoplasm and nucleus, and the proportion of CD163-positive cells was notably elevated in the CRSwNP group ($P < 0.0001$; Figure 10C).

To confirm our findings, we used qRT-PCR to examine the mRNA levels of NR4A1, CLEC4G, and CD163 in ten samples from the CRSwNP and control groups. NR4A1 levels were higher in the control group compared to the CRSwNP group ($P < 0.05$; Figure 10D), while CLEC4G ($P < 0.05$; Figure 10E) and CD163 ($P < 0.05$; Figure 10F) levels were higher in the CRSwNP group compared to the control group.

**Enrichment and Immune Correlation Analysis for Biomarkers**

According to the absolute rankings of NES, we determined the top five pathways with NES $> 0$ or NES $< 0$. The results demonstrated that all biomarkers were enriched throughout the Notch signaling, olfactory transduction, neuroactive...
ligand receptor interaction, and circadian rhythm pathways in mammals. NR4A1 is involved in starch, sucrose, retinol, and drug metabolism. CLEC4G was enriched in the calcium signaling pathway, hematopoietic cell lineage, and cytokine-cytokine receptor interaction pathway. CD163 is involved in “ABC transporters”, “steroid hormone biosynthesis”, and “complement and coagulation cascades”, among others (Figure 11A, Supplementary Table 3–5).

Additionally, the correlation between biomarkers and immune cell results showed that NR4A1, CLEC4G, and CD163 were considerably correlated with eight, four, and 12 immune cells, respectively (Figure 11B). Among them, all biomarkers were significantly correlated to effector memory CD4 T cells (NR4A1: $r=-0.68$, $P=0.0049$; CLEC4G: $r=0.63$, $P=0.0091$; CD163: $r=0.80$, $P=0.00027$), γδ T cells (NR4A1: $r=-0.78$, $P=0.00058$; CLEC4G: $r=0.51$, $P=0.042$; CD163: $r=0.59$, $P=0.018$), immature DCs (NR4A1: $r=-0.65$, $P=0.0075$; CLEC4G: $r=0.57$, $P=0.021$; CD163: $r=0.70$, $P=0.0032$), and regulatory T cells (NR4A1: $r=-0.68$, $P=0.0047$; CLEC4G: $r=0.60$, $P=0.015$; CD163: $r=0.88$, $P<2.2e-16$) (Figure 11C, Supplementary Table 6–8).
Regulation Networks and Drug Prediction

Five common miRNAs namely hsa-miR-559, hsa-miR-608, hsa-miR-623, hsa-miR-200b-5p, and hsa-miR-124-3p were obtained from the two databases (Figure 12A). Next, a ceRNA network was created, which included two mRNAs, five miRNAs, and 12 lncRNAs. Interestingly, 12 lncRNAs only regulated hsa-miR-124-3p (Figures 12B, and C). The TF-mRNA network contained three mRNAs and 17 TFs (Figure 12D). For example, GATA2 regulated all biomarkers, whereas YY1, CREB1, HINFP, and POU2F2 regulated the expression of NR4A1 and CD163.

According to the CTD databases, NR4A1 and CD163 were associated with “rhinitis”, “allergy”, and “hearing loss”. Additionally, CLEC4G was associated with “nose neoplasms” and “sinusitis” (Figure 12E). Finally, a biomarker–drug network was built, comprising one biomarker (NR4A1) and 10 drugs (Figure 12F), including ionomycin, chembl35482, and acetylcysteine.

Discussion

In our study, we detected elevated dendritic cell infiltration in CRSwNP and searched for potential biomarkers (NR4A1, CLEC4G, and CD163) associated with dendritic cells during disease progression in CRSwNP. All biomarkers were significantly correlated to effector memory CD4 T cells, γδ T cells, immature DCs, and regulatory T cells. And we constructed a dendritic cell-centred global regulatory network by analysing the biological pathways, molecular regulatory mechanisms, and relevant drugs of dc-related biomarkers in CRSwNP (Figure 13). First, we performed cluster analysis of all cells and genes detected in the three sample groups (CM, NM, and NP) after quality control. We screened myeloid cells for subcellular identification, and C3F3R was used as a neutrophil Marker gene according to the annotation results.15,16 After observing the expression of C3F3R in each subgroup of Myeloid cells, it was found that C3F3R was significantly higher in Cluster7. Therefore, Cluster7 can be identified as neutrophils. Furthermore, our results showed...
that the distribution of DCs varied considerably among the NP, NM, and CM groups. In contrast, Kim et al. found that the rate of innate immune cell infiltration was increased in the NP group. Moreover, CRSwNP tissues displayed an increased number of innate immune cells and also showed a more pronounced increase in MBP+, CD68+, and CD11c+ cells. Since CD68 and CD11c are not limited to dendritic cells (DCs), but also can be expressed by macrophages and...
**Figure 12** Prediction of regulatory networks and drugs. (A) Screening for targeted miRNAs. (B) ceRNA network. Orange: mRNA; green: miRNA; purple: lncRNA. (C) Alluvial diagram. (D) TF-mRNA network. Orange: mRNA; blue: TF. (E) Hub genes are significantly associated with ear, nose, and throat disorders. (F) Hub gene and corresponding small-molecule drug network map. Orange: hub genes; green: small-molecule therapeutics.
monocytes, the proportion of myeloid cells is significantly increased in patients with CRSwNP. It has been reported that the levels of eosinophils, dendritic cells, basophils, and neutrophils are increased in patients with CRSwNP. Myeloid cells included granulocytes, erythrocytes, megakaryocytes, and monocytes, thus consistent with our findings. That is, the level of myeloid immune cells was elevated in CRSwNP tissues compared to the other two groups, which is consistent with our findings. Notably, we showed that the immune infiltration of DC was highest in the diseased (CRSwNP) group. DC cells are specialized antigen-presenting cells in the organism that efficiently take up, process and present antigens. It has been shown that the significant increase in the number of DCs in nasal polyp tissues may be the main reason for the imbalance of the Th1/Th2 and the general predominance of the expression of Th2 cytokines in the pathological process of nasal polyps. In our study, correlation analysis among immune cells indicated a functional synergy between immature DCs and effector memory CD4+ T cells, DCs mediate the activation of T cells and receive the activation signals back from T cells. Zhang et al identified DC cells as a key determinant of the type of immune response, and specific DC cell subpopulations were isolated that could induce type 2 immunity. In our study, we screened for biomarkers of DCs, of these, both CD163 and NR4A1 were highly expressed early in CRSwNP, and CD163+ DCs was shown to correlate with the activation of naïve T cells early in the disease. Lin et al reported a substantial increase in the number of DCs in individuals with NP, which could lead to the activation of B7-1 (CD80) and B7-2 (CD86) expression through their interaction with CD40L on T cells. DC-expressed B7 molecules can activate T cells through the CD28/CTLA-4 pathway, and numerous studies have implicated a collaborative interaction involving DCs and CD4+ T cells in antigen presentation during the immune response. This process of antigen presentation operates bidirectionally, involving DCs facilitating T cell activation while concurrently receiving activation signals from T cells. Notably, T cells exhibit heightened expression of CD80/CD86 on their surfaces during this interaction and undergo reduced adhesion and phagocytosis. The interplay between DCs and T cells could potentially trigger T cell maturation and the secretion of cytokines, thereby inducing a more substantial influx of both DCs and T cells into NP tissues. Consequently, the increased quantities of DCs and T cells observed in this study may further contribute to the progression of NP.
In this study, we carried out whole transcriptome sequencing, accompanied by pairwise comparisons of the CRSwNP and control groups, to identify candidate genes that play a role in the development of NPs. Three hub genes, NR4A1, CLEC4G, and CD163, were identified using the XGBoost and SVM–RFE algorithms. NR4A1 is a member of the NR4A family of orphan nuclear receptors. These are present in various tissues and can be rapidly induced to express several stimuli. One of the most prominent functions of NR4A1 is to regulate inflammation, with its anti-inflammatory properties in various cell types, including macrophages, epithelial cells, and endothelial cells. Prior investigations have indicated that NR4A1 deficiency can intensify airway inflammation, particularly in individuals with asthma by inhibiting airway inflammation through the NF-κB/IκBα signaling pathway. NR4A1 might be an asthma biomarker and a potential prognostic factor for asthma immunotherapy. In our study, NR4A1 expression was reduced in the CRSwNP group. Pathway analysis revealed that NR4A1 was mainly involved in glucose metabolism, cell differentiation, proliferation, and apoptosis. Therefore, our results indicated that NR4A1 may act as a prognostic predictor of CRSwNP. CLEC4G, a C-type lectin protein family member, is highly expressed in monocytes. CLEC4G is the second most upregulated gene in CRSwNP and interacts with the surface glycoproteins of diverse viruses, including the spike proteins of SARS coronaviruses. Here, our study found that CLEC4G was upregulated in NP tissues, which is in agreement with previous studies. CD163 is a monocyte and macrophage transmembrane scavenger receptor protein that can be induced by glucocorticoids and IL-10. CD163 mediates the binding of monocytes to bacteria, leading to the release of inflammatory cytokines. However, the diversity, ontogeny, and functional capacity of DC subpopulations in humans are not fully understood. First, macrophages and dendritic cells have homology. They are monocyte-derived immune cells, and they share a common progenitor cell, which can be induced to differentiate into dendritic cells and macrophages in culture. In our results of pseudotime analysis, CD163 expression level was first increased and then decreased, indicating that the dendritic cell population in CRSwNP had the characteristics of M2-type macrophages at the early stage of disease progression. In addition, it has been found that circulating CD163+ DCs are a specific subpopulation of DC cells, called DC3s, which develop independently of CDC-restricted (CDP) and monocyte-restricted (cMoP) progenitor cells through a specific pathway activated by GM-CSF. DC3s can drive the activation of naive T cells. In our study, we uncovered a pivotal role played by CD163+ DCs in the progression of CRSwNP. Bourdely et al demonstrated that CD163+ DCs change through a defined pathway activated by GM-CSF. Unlike conventional DCs, CD163+ DCs activate naive T cells, which is consistent with the results obtained in this study on the synergistic effects of DC and T cells. Additionally, we examined changes in biomarker expression levels at multiple differentiation stages of DCs using pseudotime analysis after single-cell sequencing. The expression levels of both NR4A1 and CD163 initially increased and then decreased. Thus, these markers can predict the specific immune activation in sinusitis.

In our whole transcriptome sequencing enrichment analysis, we observed that all biomarkers were involved in the cytokine and cytokine receptor pathways. NR4A1 can directly inhibit NF-κB signaling and inflammatory factor expression. CD163 is a key player in the NF-κB pathway in fibroinflammatory diseases. Moreover, the NF-κB signaling pathway is critical for cytokine-induced gene expression. The CLEC4G receptor on DC surfaces can bind ligands on the surface of T cells and induce intracellular signaling via DAP12 to promote cytokine secretion, which induces initial CD4+ T cell differentiation. According to immune correlation analysis, the three biomarkers were significantly associated with T cells, and previous studies have shown that NR4A1 can be rapidly induced by signaling by T cell antigen receptors and removal of CD163 leads to massive infiltration of activated T cells. Furthermore, CLEC4G triggers CD8+ T cell tolerance and CD4+ T cell differentiation. In addition, CTD database analysis revealed that the biomarkers CD163 and NR4A1 were associated with rhinitis, whereas CLEC4G was associated with sinusitis. In our constructed biomarker–drug network, we found one biomarker (NR4A1) and 10 drugs, including ionomycin, chembl35482, and acetylcysteine. Acetylcysteine has been used in the antioxidative stress treatment of CRSwNP. Nicotine is an exogenous nicotinic acetylcholine receptor agonist that suppresses proinflammatory cytokines release by triggering cholinergic anti-inflammatory pathways. It is used clinically to treat ulcerative colitis and combat inflammation. According to recent studies, nicotine is a potential new CRSwNP treatment for patients with severe SARS-CoV-2.

Our study presented reliable DC biomarkers and their associated signaling pathways, and the identified DEGs represent a valuable resource. By combining scRNA-seq and whole transcriptome sequencing to analyze immune and non-immune cells from mucosal tissue samples extracted from nasal polyps, nasal mucosa tissues neighboring nasal
polyps, and nasal mucosa tissue samples from a control group, and identify DCs as a key immune cell population in CRSwNP and three hub genes as biomarkers of CRSwNP. Furthermore, the constructed ceRNA, TF-mRNA, and biomarker–drug networks provided new targets for CRSwNP treatment.

Limitations
Our study had some limitations. We identified 3 hub genes related to DCs in CRSwNP, the precise roles of these hub genes in the pathogenesis of CRSwNP need to be further elucidated. Therefore, extensive animal experiments are required to explore these functions. In addition, we need more research to confirm our findings by expanding the sample size.

Conclusion
We identified DCs as a key immune cell population in CRSwNP and three hub genes (NR4A1, CLEC4G and CD163) as biomarkers of CRSwNP by combining scRNA-seq and whole transcriptome sequencing. CD163 and NR4a1 were highly expressed in the early stage of CRSwNP and were involved in the activation of T cells. Furthermore, the constructed ceRNA, TF-mRNA, and biomarker–drug networks provided new targets for CRSwNP treatment. This study provides potential therapeutic agents for the clinical treatment of CRSwNP based on key genes.

Abbreviations
AUC, area under the curve; CRS, chronic rhinosinusitis; CRSwNP, chronic rhinosinusitis accompanied by nasal polyps; NP group, nasal polyps; NM group, nasal mucosa tissue samples adjacent to nasal polyps from the CRSwNP group; CM group, nasal mucosa tissue samples from the control group; DCs, Dendritic cells; DCRGs, DC-related genes; ceRNA, endogenous RNA; GEO, Gene Expression Omnibus; GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; PCA, principal component analysis; qRT-PCR, quantitative real-time reverse transcription-polymerase chain reaction; scRNA, single cell; SVM–RFE RNA, Support vector machine-recursive feature elimination; TF-mRNA, transcription factor; UMAP, Uniform manifold approximation and projection.

Ethics Approval and Informed Consent
The Beijing Chaoyang Hospital's Ethics Committee’s Institutional Review Board (No. 2020-3-16-18) gave its approval to this work. All experiments involving human participants adhered to the institutional ethical guidelines and followed the principles of the Declaration of Helsinki. Before participating in the study, all participants involved provided written informed permission.

Data Sharing Statement
The unique inputs included in the study are outlined in the article/supplementary material. Further questions or concerns may be directed to the corresponding author.

Acknowledgments
We appreciate Dr. Lu’s insightful advice and constructive comments regarding our study.

Funding
This research received funding support from the National Natural Science Foundation of China (82071068).

Disclosure
The authors report no conflicts of interest in this work.

References


