How Cytokines Co-occur across Asthma Patients:
From Bipartite Network Analysis to a Molecular-Based Classification

Suresh K. Bhavnani$^{1,2,4}$ PhD, Sundar Victor$^1$ MS, William J. Calhoun$^{1,3}$ MD, William W. Busse$^5$
MD, Eugene Bleecker$^6$ MD, Mario Castro$^7$ MD, Hyunsu Ju$^{1,2}$ PhD, and Allan R. Brasier$^{1,3}$ MD
$^1$Instit. for Translational Sciences, $^2$Preventive Medicine & Community Health, $^3$Dept. of Medicine, UTMB; $^4$School of BMI, Univ. of Texas, Houston, TX; $^5$Dept. of Medicine, Univ. of Wisconsin, Madison, WI; $^6$Wake Forest Univ.
School of Medicine, Winston-Salem, NC; $^7$Dept. of Medicine, Washington Univ. in St. Louis, St. Louis, MO

Abstract

Asthmatic patients are currently classified as either severe or non-severe primarily based on their response to glucocorticoids. However, because this classification is based on a post-hoc analysis of treatment response, it does not inform the rational staging of disease or therapy. Recent studies in other diseases suggest that a classification which includes molecular information could lead to more accurate diagnoses and prediction of treatment response. We therefore measured cytokine values in lung fluids obtained from 83 asthma patients, and used bipartite network visualizations with associated quantitative measures to analyze the co-occurrence of cytokines across patients. The analysis helped to identify three clusters of patients which had a complex but understandable interaction with three clusters of cytokines. Furthermore, while the patient clusters were significantly different based on key pulmonary functions, they had no significant relationship to the current classification of asthma patients. These results suggest the need to define a molecular-based classification of asthma patients, which should improve the diagnosis and treatment of this disease.

Introduction

Asthma is a chronic inflammatory disease of the airways, characterized by recurrent airflow obstruction and hyperactivity to nonspecific stimuli. Although many asthma patients respond well to inhaled glucocorticoid therapy, a subset of patients are unresponsive resulting in high rates of morbidity and mortality. Treatment for this subset accounts for more than 40% of the total cost of asthma treatment. Unfortunately, relatively little is known about which patients will have poor outcomes. For example, although asthma patients are currently classified as severe or non-severe based on their therapeutic response to glucocorticoids, this course-grained clinical classification does not explain the varying degrees of lung function compromise, airway hyper-reactivity, gastro-esophageal reflux, and COPD in patients currently diagnosed with severe asthma.

Recent developments in molecular biology and powerful analytical methods such as network analysis provide new opportunities to shift our understanding of diseases from a morphological (based on clinical and histological findings) to a molecular basis. For example, gene expression analyses have been shown to improve prediction of treatment response in several diseases such as breast cancer.

Because cytokines like Eotaxin are known to mediate airway inflammation, we hypothesized that patterns in how cytokines co-occurred across patients could provide insights for developing a new molecular-based classification of asthma patients. Such a classification, based on effector proteins found in lung fluids, could enable more accurate prediction of disease progression and therapeutic response. In contrast to our previous analysis which was based on an a priori classification of patients, here we focus on the use of network visualizations and analyses in a data-driven approach without those assumptions.

We begin by describing how we assembled a dataset of patients and their cytokine profiles, why and how we represented it using networks, and how we analyzed the networks using visualizations and quantitative measures. We then discuss how the bipartite network analysis revealed complex co-occurrence patterns of cytokine across patients, and how those patterns relate to key attributes of pulmonary function. We conclude by discussing the need to define a molecular-based classification of chronic asthma patients, and the utility of network analyses to understand complex relationships.

Method

Our research began with the question: How do cytokines implicated in asthma co-occur across patients? To address our research question, we made critical decisions regarding data selection, data representation and data analysis as discussed below:

Data Selection. Our study was based on a secondary analysis of cytokine profiles collected in a consortium-wide study. Levels for 25 cytokine were measured from lung tissue washings (referred to as bronchoalveolar lavage samples) obtained from 40 severe, and 43 non-severe asthma patients. The classification of patients was made according to the consensus definition of the American Thoracic Society, and the two groups were age and sex matched. The dataset included 11 baseline and 2
dynamic pulmonary functions relevant to the current analysis as shown in Table 1. Because 50% of values in 7 cytokines (IL-1b, IL-7, IL-10, IL-12, IL-13, IFN-γ, and GM-CSF) had undetectably low values, they were removed from the dataset, resulting in a total of 18 cytokines (see our earlier publication for details about the data collection and inclusion criteria).

**Data Representation.** Networks are increasingly being used to analyze a wide range of molecular phenomena, such as gene and protein-protein interactions, and to assess their relationships to diseases, symptoms, and syndromes. A network consists of nodes and edges; nodes represent one or more types of entities (e.g., patients or cytokines), and edges between the nodes represent a specific relationship between the entities. Figure 1 shows a bipartite network (where edges exist only between different types of entities) of patients and cytokines.

**Node diameter** was used to represent the sum of the edge weights connected to it. This enabled a rapid visual inspection to determine for example, which patients have overall high aggregate cytokine values, and how such patients relate to the rest of the network. In addition, the **node color** was used to represent asthma severity (red for severe, and black for non-severe), which enabled us to analyze how the patterns in the overall network related to the existing classification of asthma (this figure is not shown).

**Edge weights** in the network were used to represent the strength of the cytokine values for each patient-cytokine pair. Because the cytokines had different ranges, we used the min-max normalization method (cytokine value – min / max – min) to map each cytokine value to range from 0-1. This method helped to preserve the relative distances between the values, and enabled a consistent method to compare the different cytokine values. As shown in Figure 1, the edge thicknesses were drawn to be proportional to these normalized cytokine values.

**Global patterns** in a network can be rapidly visualized and analyzed using a set of network algorithms. For example, Figure 1 shows how the Kamada-Kawai layout algorithm helps to visualize the relationship between patients and cytokines. The algorithm pulls together nodes that are strongly connected, and pushes apart nodes that are not. This algorithm is fast but approximate and well-suited for medium sized networks consisting of between 100-1000 nodes. As shown, the result is that nodes with a similar pattern of connections (e.g., Eotaxin and IL-4 in the lower right hand side of Figure 1) are placed close to each other. The networks were created using Pajek (version 1.23).

A key advantage of the above network representation is the simultaneous visualization of multiple raw values (patient-cytokine associations, cytokine values, patient attributes), aggregated values (sum of cytokine values), and emergent global patterns (clusters) in a uniform visual representation. Such a representation enables the rapid generation of hypotheses based on complex multivariate relationships.

**Data Analysis.** The insights derived from the network visualizations were quantitatively analyzed using three methods. (1) Because the network layout suggested the presence of distinct clusters, we used the agglomerative hierarchical clustering method to identify the boundaries of the clusters, and a heat map to inspect the profiles of specific cytokines and patients. The clustering was done using the standard Jaccard dissimilarity measure with the Ward linkage function. Cluster boundaries were determined based on natural breaks in the patient and cytokine dendrograms. (2) To analyze the relationship between asthma severity and the patient clusters, we used the chi-square test of independence. To analyze the overall significance of 13 pulmonary functions, we used the one-way, two-tailed Kruskal-Wallis test (non-parametric ANOVA) to address the skewed values, and the false discovery rate (FDR) procedure to correct for multiple comparisons. (3) To analyze the significance between each pair of clusters for the above patient variables, we used the Dunn’s test procedure.

**Results**

The bipartite network visualization and quantitative analysis revealed distinct patient clusters, and cytokine clusters. For each set of clusters we describe the results of the visual analysis, the cluster analysis, and their significance to clinical attributes and molecular processes.

**Patient Clusters**

**Visual Analysis.** As shown in Figure 1, the visual analysis helped to identify three clusters of patients based on their cytokine profiles: (a) **Patient-Cluster-1** (shown in the lower right hand corner of Figure 1) had medium to high levels of the Eotaxin and IL-4. However, they had relatively lower values for the rest of the cytokines as shown by their relatively small diameters. (b) **Patient-Cluster-2** (shown in the center of the network) had high values of Eotaxin and IL-4, but also high values for another set of six cytokines (IL-5, IFN-γ, MIP1a, MIG, IL-17, MIP-1b) shown in the center of the network. The higher cytokine values result in relatively larger node diameters compared to Cluster-1. (c) **Patient-Cluster-3** has overall lower
values of many cytokines resulting in them being scattered along the top periphery of the network. The overall lower levels of most cytokines result in relatively smaller node diameters.

Cluster Analysis. As shown by the patient dendrogram on the vertical axis of Figure 2, the agglomerative hierarchical clustering identified the boundaries of the visual clusters in the network. Furthermore, while Patient-Cluster-1 and Patient-Cluster-2 were intuitively clear from the network, Patient-Cluster-3 was identified as a distinct cluster in the dendrogram because its members have a pattern of similarly low cytokine levels.

Relationship to Clinical Variables. To infer the meaning of the three patient clusters, we analyzed the relationship between each identified cluster to asthma severity, and pulmonary function.

Asthma Severity. As discussed earlier, patients are currently classified as severe or non-severe. An inspection of the network where patient nodes were colored based on severity, showed no visual pattern – there appeared to be an even number of both types of severity in each cluster. The chi-square analysis verified this visual result, which showed no significant association in asthma severity between the three patient clusters ($\chi^2(2,N=83)=0.9298, p=0.628$). This suggests that a classification of patients based on cytokine profiles does not match the current classification of asthma based on severity.

Pulmonary Function. As shown in Table 1, the Kruskal-Wallis test revealed that 10 out of 13 pulmonary function measures were significantly different across the clusters. The pair-wise analysis revealed that Patient-Cluster-3 had the highest number (9 out of 13) of significantly different pulmonary functions when compared to Patient-Cluster-1, and slightly fewer differences (5 out of 13) when compared to Patient-Cluster-2. In contrast, Patient-Cluster-1 and Patient-Cluster-2 had only one pulmonary function that was significantly different (results of which and how specific functions were significantly different in the
pair-wise analysis are beyond the scope of the current paper and is a focus of our current analysis).

**Cytokine Clusters**

**Visual Analysis.** The bipartite network visualization also revealed three cytokines clusters, which have a complex relationship to the patient clusters. (a) **Cytokine-Cluster-1** (in the lower right hand side of the network) consisting of Eotaxin and IL-4 contain cytokines that are pushed together because many patients from Patient-Cluster-1 and -2 have high values of those two cytokines. Their resulting large diameters suggest that they are over-represented in patients compared to the other cytokines. This observation is also salient by the many red cells (representing high values) in the last two columns (representing Eotaxin and IL-4) of the heat map in Figure 2. (b) **Cytokine-Cluster-2** consisting of six cytokines (mentioned earlier) which are pushed together because they have high values of mainly Patient-Cluster-2. Unlike Cytokine-Cluster-1, they have high values for only one patient cluster, and therefore have smaller diameters. (c) **Cytokine-Cluster-3** consisting of the remaining cytokines scattered on the left and right hand side of the network have overall lower values across all patients, and therefore have the smallest diameters in the network.

**Cluster Analysis.** As shown by the cytokine dendrogram on the horizontal axis of Figure 2, the agglomerative hierarchical clustering identified the boundaries of the visual clusters in the network. While Cytokine-Cluster-1 and Cytokine-Cluster-2 are intuitively clear from the network, Cytokine-Cluster-3 is identified as a distinct cluster in the dendrogram because it has a pattern of similarly weak levels with patients. This observation is salient by the large number of green cells (representing low values) for this cluster in the heat map in Figure 2.

**Discussion**

The results suggest that cytokine values can indeed separate patients into distinct clusters. While this result was sufficient on its own for insights to classify asthma patients, the network analysis also helped to identify cytokine clusters which enabled us to infer biological meaning about the patient clusters.

The frequent co-occurrence of Eotaxin and IL-4 (Cytokine-Cluster-1) is congruent with a known sequence of molecular changes in asthma patients who often have a T-helper-2 (TH2) lymphocyte-skewed immune response. This response results in the secretion of IL-4, which in turn induces Eotaxin production by bronchial epithelial cells. The resulting downstream actions include the activation and recruitment of tissue-resident eosinophils, a hallmark of early stage asthma. The presence of Eotaxin and IL-4 in lung fluids therefore represents important sub-stages of a complex molecular pathway in asthma, which explains their frequent co-occurrence in the network.

To understand the biological significance for cytokines in Cytokine-Cluster-2 (IL-5, IFN-γ, MIP1a, MIG, IL-17, and MIP-1β), we entered its members into the Ingenuity Pathway Analysis (IPA) application. The results from IPA suggest that the frequent co-occurrence of these cytokines is regulated by the innate inflammatory nuclear factor-
κB pathway. NF-κB is a potent pro-inflammatory transcription factor that activates expression of cytokine networks. Furthermore, persistent NF-κB activation has been linked to uncontrolled/acute exacerbations of asthma. The frequent co-occurrence of this set of cytokines therefore implies the presence of a distinctly different pro-inflammatory state compared to the IL-4–Eotaxin process.

The above cytokine clusters, along with pulmonary functions of the patients, provide a biological explanation for the patient clusters. The strong relationship of Patient-Cluster-1 to Cytokine-Cluster-1 suggests that patients in this cluster have disease primarily driven by T\(_{h2}\) inflammation. In contrast, Patient-Cluster-2 has a strong relationship to both Cytokine-Clusters-1 and -2. This result implies that patients in Patient-Cluster-2 have a component of activated innate inflammatory pathways. Further evidence for this inference of state-based clusters is provided by differences in pulmonary function across the clusters: Patient-Cluster-3 which has the lowest cytokine values for both of the above cytokine clusters, also has the largest number of significant differences in obstructive airway disease parameters in pulmonary function testing, and lowest airway reactivity response to methacholine compared to Patient-Clusters-1 and -2. This implies that Patient-Cluster-3 represents a subgroup of asthmatics with preserved pulmonary function and greatest response to albuterol without active inflammation. The network analysis of patients and cytokines therefore implies a state-based classification of asthma patients informed by underlying molecular processes.

The limitation of our study is that we analyzed only one dataset, and our future research will attempt to replicate the results in a similar dataset. However, the current results suggest that asthma patients can be meaningfully classified using molecular markers such as cytokines.

**Conclusion and Future Research**

Cytokines control key processes in asthma including immune activation and T lymphocyte skewing. However, little work has been done to investigate whether and how cytokines could help to classify patients. By using bipartite network visualizations without *a priori* assumptions of patient classes, combined with appropriate quantitative methods suggested by the patterns in the network, we arrived at a new state-based understanding of asthma. It is important to note that the bipartite network could have revealed co-occurrence patterns without the presence of distinct clusters, prompting us to use other methods to quantify the patterns. Therefore, we believe that bipartite networks provide an important first step to identify the nature of co-occurrence in molecular data, which then guides the use of appropriate quantitative methods. The bipartite network representation also allowed us to overlay multiple raw and aggregated variables onto the same visualization enabling us to analyze complex multivariate relationships.

In our future research, we plan to analyze other multivariate relationships in the network, and validate the state-based classification of asthma in similar datasets. This should lead in the future to a classification of asthma patients that is based on underlying biological processes, resulting in more effective diagnosis and treatment.

**Acknowledgements**

This work was supported by NIH grants 1U54RR02614 UTMB CTS(A(RB), AI062885(ARB), NHLBI contract BAA-HL-02-04(ARB), HL69130 US SARP(WJC), and HL69149 (MC). We thank H. Spratt, M. Sinha, A. Ganesan, and D. Bostick for their suggestions.

**References**