Hierarchical clustering for biomarker discovery in immune-mediated neurological disorders

De Martino Maria[1], Fabris Martina[2], Fiorino Roberta[2], Isola Miriam[1]

1 Division of Medical Statistic, Department of Medicine - University of Udine. Udine, Italy
2 Department of Laboratory Medicine, University of Udine, ASUFC, Udine, Italy

Introduction

The imbalance in cytokine regulation contributes to the onset of various conditions, including immune-related neurological disorders. These diseases involve an inflammatory and demyelinating response targeted at antigens located in the central and peripheral nervous systems. Cytokines are highly effective biomarkers for detailing the inflammatory status of these disorders. It is interesting to explore whether different biomarkers can reveal unique cytokine expression profiles that distinguish a specific disease or identify subgroups of patients with diverse phenotypes within the same disease. Hierarchical clustering (HCA) is a machine learning technique that forms groups of objects, ensuring that objects within the same group are similar to each other while differing from objects in other groups. In this context, it can play an interesting role in subgrouping patients.

Objective

Use hierarchical clustering to identify subgroups of patients with diverse phenotypes within immune-related neurological disorders.

Methods

The study includes 103 patients with immune-related neurological diseases who underwent cerebrospinal fluid sampling. Demographical informations as age and sex were collected, while values of cytokines were extracted from the samples. The biomarkers chosen were: IL-1b, IL-8,TNF-a, CXCL10, IFNg, IL-10, IL-2ra, IL-6, BLC/CXCL13. Before starting the analysis, missing data were replaced using K-Nearest Neighbors imputation. Furthermore a Density-Based Spatial Clustering of Applications with Noise (DBSCAN) model was applied for noise points removal. HCA was then implemented for the creation of six different clusters. Furthermore four expression levels of cytokines (low, intermediate, high and very high) were determined using the 25^{th}, 50^{th} and 75^{th} percentile, used for describing the various clusters.

Results

DBSCAN methods led to the removal of 7 patients from the dataset, leaving with a dataset of 96 patients. HCA defined 6 clusters of different sizes: 29 patients in cluster 1, 12 in cluster 2, 2 in cluster 3, 24 in cluster 4, 29 in cluster 5 and 10 in cluster 6. Using the expression levels defined before, cluster 1 was intermediate inflammatory, while cluster 6 was high level inflammatory.

Conclusion

HCA results to be a useful tool to better help the understanding of inflammatory pattern in immune-related neurological diseases, leading to the possibility of specific therapies depending on the cytokine levels.

Bibliography

1. Ward J. H. et al. Hierarchical Grouping to Optimize an Objective Function. Journal of the American Statistical Association, 1963; 58(301), 236–244.

Table 1. Clusters obtained from HCA: (-) low; (+) intermediate; (++) high; (+++) very high

CLUSTER	IL-1B	IL-8	TNF-A	CXCL10	IFNG	IL-10	IL-2RA	IL-6	BLC/CXL13
1 (N=29)	++	+	+	+	++	+	+	+	+
2 (N=12)	+++	++	++	++	+	++	++	+	++
3 (N=2)	+	++	++	++	+++	+++	++	+	-
4 (N=24)	++	+++	++	+	++	+	+	+++	++
5 (N=29)	+	++	+	++	++	++	++	+	+++
6 (N=10)	++	+++	+++	+++	+++	+++	+++	+++	+++