Clustering of Exposome and Lifestyle Data to Support Differential Expression Analysis of Circulating miRNAs in Multiple Sclerosis

Ercoli Vittoria[(1)], De Giuseppe Rachele[(2)], Bertoli Gloria[(3)], Alemayohu Mulubirhan Assefa[(1)], Al-Naqeb Ghanya[(2)], Ballante Elena[(4)], Bergamaschi Roberto[(4)], Camboni Tania[(5)], Ceccarani Camilla[(5)], Ceriani Chiara[(3)], Colombo Elena[(4)], Consolandi Clarissa[(5)], Costabile Francesca[(6)], De Cesari Stefano[(6)], Filosa Asia[(1)], Gallivanone Francesca[(3)], Galuzzi Bruno Giovanni[(3)], Gastaldi Matteo[(4)], Gervasoni Clarissa[(5)], Gugiatti Simona[(2)], Itri Teresa[(3)], Kalmpourtzidou Aliki[(2)], Landi Tony[(6)], Lanzotti Aurora[(3)], Lo Dico Alessia[(3)], Loperfido Federica[(2)], Maccarini Beatrice[(2)], Mazza Antonio[(7)], Montomoli Cristina[(1)], Oddone Enrico[(8)], Pellizzer Chiara[(3)], Pernetti Roberta[(8)], Porro Danilo[(3)], Scarabotto Anna[(4)], Sellaro Francesca[(8)], Severgnini Marco[(5)], Summa Donato[(7)], Tagliaferri Sofia[(2)], Tani Flaminia[(3)], Torriani Camilla[(1)], Vivone Gemine[(7)], Cena Hellas[(2, 9)], Tavazzi Eleonora[(4)], Monti Maria Cristina[(1)]

(1) *Biostatistics and Clinical Epidemiology Unit, Department of Public Health, Experimental and Forensic Medicine, University of Pavia, Pavia*
(2) *Laboratory of Dietetics and Clinical Nutrition, Department of Public Health, Experimental and Forensic Medicine, University of Pavia, Pavia*
(3) *Istituto di Bioimmagini e Sistemi Biologici Complessi - CNR, Segrate (MI)*
(4) *Multiple Sclerosis Research Center, IRCCS Mondino Foundation, Pavia*
(5) *Istituto di Tecnologie Biomediche, CNR*
(6) *Istituto di Scienze dell'Atmosfera e del Clima, CNR*
(7) *Istituto di Metodologie per l'Analisi Ambientale, CNR*
(8) *Occupational Medicine Unit, Department of Public Health, Experimental and Forensic Medicine, University of Pavia, Pavia*
(9) *Clinical Nutrition Unit, Department of General Medicine, Istituti Clinici Scientifici Maugeri IRCCS, Pavia*

CORRESPONDING AUTHOR: *Ercoli Vittoria, ercolivittoria@gmail.com*

INTRODUCTION

MicroRNAs (miRNAs) are small non-coding RNAs that regulate gene expression post-transcriptionally and are influenced by various metabolic and environmental factors[1]. The functional exposome concept can be seen as a new strategy to study the effect of the environment on health and is increasingly studied to understand its role in Multiple Sclerosis (MS) development[2]: it describes the harmful biochemical and metabolic changes (internal exposome) that occur in our body due to the totality of different environmental exposures throughout the life course, ultimately leading to adverse health effects and premature deaths.

OBJECTIVE

Variations in circulating miRNAs, as biomarkers of inflammation and oxidative stress, may identify subgroups of MS patients at risk. Cluster and study the characteristics of those groups and prospect different pattern of miRNAs, that may reflect distinct health status.

METHODS

A cohort of 139 people with Multiple Sclerosis (pwMS) was evaluated with detailed external exposome and lifestyle (air quality, urbanization, nutritional and occupational status) and internal exposome data (microbiome, oxidative stress and inflammation biomarkers).

Differential expression levels of five circulating miRNAs (miR-30, miR-146, miR-330, miR-574, and miR-664) have been measured and normalized with respect to an endogenous control in all blood samples, for each miRNA; they were obtained from a reduced subset of participants (sample sizes ranging from 15 to 25 subjects per miRNA).

Principal Component Analysis (PCA)[3] was used to reduce dimensionality of datasets and identify key patterns among the external exposome and lifestyle data. The first two principal components, which account for the most significant portion of variance, were selected based on parallel analysis of eigenvalues. Elbow Method was used to validate the optimal number of cluster and K-means clustering was then performed using these components.

DOI: 10.54103/2282-0930/29376

To assess the miRNA expression differences among cluster membership, we performed Levene's test for homogeneity of variance followed by Kruskal–Wallis non-parametric tests. When appropriate, Dunn's test with Bonferroni correction was used for post hoc comparisons.

RESULTS

PCA identified two meaningful components that captured the primary axes of variation in lifestyle and nutritional profiles (Figure 1A). PC1 was primarily characterized by negative loadings on EDSS while the PC2 by positive; regarding anthropometric variables (e.g., BMI, waist-related measures) both PCs are characterized by negative loadings. Whereas PC2 is correlated negatively with quality of life (MSQOL-29 both physical and mental) in contrast, PC1 showed a positive contribution of those variables.

Based on the Elbow Method, three clusters were identified as optimal. K-means clustering was then applied to the first two PCA-derived components, resulting in the classification illustrated in Figure 1B. Three distinct clusters emerged: Cluster 1 (blue), positioned predominantly in the upper-left quadrant, Cluster 2 (yellow), mainly located on the right-hand side of the plot, and Cluster 3 (grey), concentrated in the lower-left area. These clusters differ along both PC1 and PC2 axes, suggesting heterogeneity in the underlying data structure. The clustering indicates that individuals were grouped based on shared patterns in anthropometric, clinical, and lifestyle variables.

No statistically significant differences in expression were found for any of the five miRNAs across clusters. Levene's tests confirmed the homogeneity of variances in all comparisons ($p > 0.1$) and, since miRNA levels were not normally distributed, the Kruskal–Wallis test was used. Kruskal–Wallis tests for miR-30, miR-146, miR-330, miR-574, and miR-664 all yielded non-significant p-values.

CONCLUSIONS

PCA and clustering analysis should be considered as valuable tools to summarize complex lifestyle variables and explore their interrelationships in MS research. This exploratory analysis using unsupervised clustering of exposome-lifestyle data did not reveal significant associations with expression of selected circulating miRNAs in pwMS, but further comprehensive miRNA profiling on the full sample is warranted to validate the negative results obtained or change the proof of concept.

REFERENCES

1. Y. Gao, D. Han, e J. Feng, «MicroRNA in multiple sclerosis», Clin. Chim. Acta, vol. 516, pp. 92–99, mag. 2021, doi: 10.1016/j.cca.2021.01.020.

2. M. C. Monti et al., «Exposome, oxidative stress and inflammation in multiple sclerosis: the EXPOSITION study protocol», Eur. J. Public Health, vol. 34, fasc. Suppl 3, p. ckae144.1431, ott. 2024, doi: 10.1093/eurpub/ckae144.1431.

3. A. Simanjuntak e M. S. Hasibuan, «Application of PCA and K-Means Clustering Methods to Identify Diabetes Mellitus Patient Groups Based on Risk Factors», Prisma Sains J. Pengkaj. Ilmu Dan Pembelajaran Mat. Dan IPA IKIP Mataram, vol. 11, fasc. 4, p. 1002, ott. 2023, doi: 10.33394/j-ps.v11i4.9263.

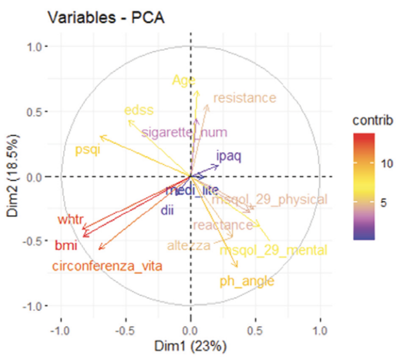

Figure 1A: Variables contributions to Principal Components (PCA) 1 and 2

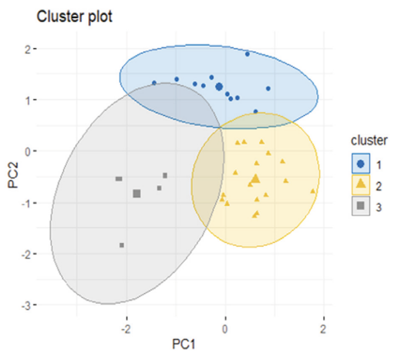

Figure 1B: Cluster assignment based on PCA derived components

DOI: 10.54103/2282-0930/29376