# SUPPLEMENTARY MATERIAL

## Supplementary information

## METHODS

### Statistical Methods- Study variables

The variables included in the analysis were selected as per clinical criteria and statistical criteria (variables with many missing values were excluded). There was no control group. Therefore, it is a descriptive analysis in which no comparison between the different patient groups will be presented as per their treatment response but based on their pre-treatment baseline characteristics. Of the 345 evaluable patients from the FENOMA study, 89 were not included in the cluster analysis due to missing data. Thus, 256 patients (74%) served for cluster analysis.

Descriptive statistics were calculated for demographic and clinical characteristics. Continuous variables are presented as median and quartile 1 (Q1) and quartile 3 (Q3). Categorical variables are reported by using frequency and percentage.

For statistical comparisons, the following tests were used, as applicable: Kruskal-Wallis test, Chi-square test, Fisher's exact test, Wilcoxon signed-rank test, and t-test. A 2-tailed P value of less than 0.05 was used to denote statistically significant differences. All statistical analyses were performed with SAS® version 9.4 and SAS® Enterprise Guide V7.15 (SAS Institute, Cary, NC).

*Variable correlation analysis*

In the case of highly correlated variable groups, a representative of each group is selected, as the information that these variables would provide to the analysis would be redundant, and it may

 
doi: 10.18176/jiaci.0731

hinder the clustering process. Of these primary variables, only those with at least 85% of available data were included **(Table 1).**

*Search for the optimal number of clusters*

Variance ratios that were explained for a wide range number of clusters were analyzed to find the optimal number of clusters in the calculation using the elbow method. **Supplementary Figure 1** shows the candidates found. Based on the explained variance progression, the optimal minimum number of clusters was 4 clusters.

*Analysis of the result for the optimal number of clusters*

Following the number of clusters found in the previous step, a clustering analysis was performed to obtain an identification cluster number that was assigned to each patient. **Supplementary Table 2** summarizes the number of patients that set each group and additional analysis data. Two main groups formed by 141 and 96 patients were created, while there were also two small groups with 12 and 7 patients. The latter groups were unrepresentative due to their small sizes.

To overcome the high number of variables analysis, canonical discriminant analysis and principal components analysis were performed. With the information on how those vectors were related to the original variables, the most representative variables (demographic, clinical, pulmonary function, biological markers) were extracted when classifying the patients in clusters to create a profile of the patients that formed each group. **Supplementary Table 3** shows the mean values of the canonical vectors for each of the clusters. The first canonical vector (Can1) significantly differentiated cluster 4 from the other clusters, the second vector (Can2) significantly differentiated

 
doi: 10.18176/jiaci.0731

cluster 3 from the other clusters, and the third vector (Can3) had different values for clusters 1 and 2, which helped to differentiate these two groups. The three canonical vectors (Can1, Can2, and Can3) are represented in **Supplementary Figure 2** in pairs.

*Correlation of cluster profiles with post-treatment parameters*

After the cluster analysis, the correlations between the groups of patients (clusters) and some post-treatment parameters were examined through a multivariate analysis. Specifically, the aim was to analyze if there were any differences in response to treatment between the main clusters; therefore, the two smaller ones were excluded from this study



doi: 10.18176/jiaci.0731