Biostatistics statement

The SPSS and GraphPad statistical software programs were used for the analysis. Non-parametric tests were used for statistical analysis because of the skewed distribution of FC levels. In particular, comparison between groups was performed by Mann-Whitney (2 groups) or Kruskal-Wallis test (>2 groups). Values of FC from the same patient were compared by Wilcoxon test. Categorical variables were compared by chi-squared test. Spearman's r-test was used to assess correlations of FC with other variables.

Univariable and multivariable Cox regression analysis was performed to evaluate whether FC could serve as independent predictor for disease flare. We accepted the hypothesis that the predictive accuracy of FC for disease flare gradually decreases over time so a 6-month cut-off time-point was tested as described above. Proportional hazard assumptions of other predictive factors were graphically evaluated using log-minus-log plots. The predictive value of FC was adjusted in multivariable Cox regression for all variables achieving a two-sided $P<0.20$ in univariable Cox regression. Hazard ratios (HR) were presented per 100 μg/g increment for FC levels.

Receiver operating characteristic (ROC) curve analysis was used to determine specificity, sensitivity and optimal threshold values of FC. The accuracy of FC was evaluated using the area under the curve (AUC) of the ROC and was defined as follows: poor 0.6–0.7; fair 0.7–0.8; good 0.8–0.9; excellent 0.9–1.0. In all cases, an alpha level of <0.05 was considered to be significant.

The statistical methods of the study were conducted and reviewed by Spyros I Siakavellas, a member of our research group with a MSc in Biostatistics, with considerable experience as a biomedical statistician.