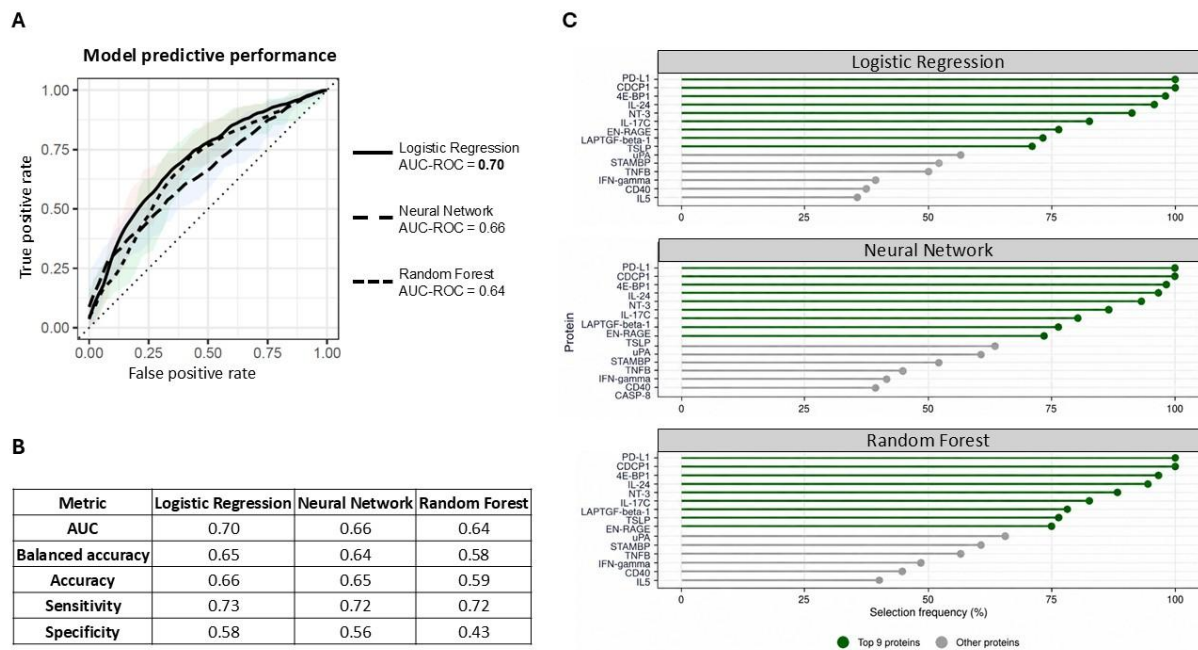




Supplementary Figure S7. Comparison of machine-learning classifiers in b/tsDMARD-naïve RA patients. – (A) Receiver operating characteristic (ROC) curves obtained from repeated nested cross-validation for logistic regression, neural network, and random forest classifiers. Starting from 50 pre-specified candidate proteins derived from the *ex vivo* JAKi- and TNFi-modulated signatures, feature selection and panel-size optimization were performed exclusively within the training partitions. Models were evaluated using 20 repetitions of outer 10-fold cross-validation with an inner 3-fold loop. Curves represent the mean out-of-fold ROC across repetitions, with shaded areas indicating the empirical 2.5th–97.5th percentile range. Logistic regression showed the highest mean AUC (0.70), followed by neural network (0.66) and random forest (0.64). (B) Summary of model performance, including AUC, balanced accuracy, sensitivity, specificity, and overall accuracy. Logistic regression showed the highest overall performance across the evaluated metrics. (C) Protein selection frequencies across repeated nested cross-validation for each classifier. Proteins highlighted in green correspond

to the nine most frequently selected features within each model, whereas the remaining proteins are shown in grey.