

Convergent Random Forest Predictor: A new method for designing predictors of drug response. Predicting response to anti-TNFs in Rheumatoid Arthritis using whole blood gene expression.

Prediction of patients' response to therapy using molecular biomarkers is one of the goals of transcript profiling of human tissues. Here we present a powerful analysis approach, the Convergent Random Forest (CRF) method, for the identification of highly predictive biomarkers. The aim is to select from genome-wide expression data a small number of biomarkers that could be developed into a simple and robust diagnostic tool. Due to the large number of transcripts, relatively limited number of samples, and high variability of data, identification of predictive biomarkers is a challenge for data analysis. Additionally, it is clear that many genes are responsible for drug response differences given the complexity of the biological processes. The driving principle of our approach is to limit the noise that contributes to the candidate predictor genes and to select the smallest number of non-redundant and most predictive genes. Our method combines the Random Forest classifier and gene expression clustering to rank and select a small number of most predictive genes. Using our method we have analyzed transcript profiles of whole blood collected before TNF-blockade treatment from Rheumatoid Arthritis patients. We have identified 8 transcripts that predict response to therapy with 91% specificity and 87% sensitivity in cross-validated sets.