

miRNA-BASED PROGNOSTIC PREDICTOR FOR OVARIAN TUMORS USING MACHINE LEARNING

Alexandre Chiavegatto Filho, Mariana Boroni, Cristiane Esteves Teixeira

Brazilian National Cancer Institute (INCA)

Abstract

Ovarian cancer is one of the neoplasms with the highest incidence among women worldwide, with significantly high mortality. The vast majority of patients are diagnosed in advanced stages of the disease since early stages present unspecific symptoms and inaccurate diagnoses. Currently, there are not many biomarkers available clinically to help diagnose or predict its prognosis because of contradicting diagnostic accuracy. In this sense, the application of multi-omic integration approaches combined with machine learning techniques is promising, not only to better understand cancer prognosis but to identify effective prognostic biomarkers related to ovarian cancer. Therefore, our goal is to build a predictor of prognosis for patients diagnosed with ovarian cancer, as well as to identify new biomarkers. Patients from diagnosed with ovarian cancer obtained from the database of The Cancer Genome Atlas (TCGA) project were classified into two groups according to their prognosis. Patients with < 3 years of survival and Dead status were allocated as the group with poor prognosis and patients with = 3 years were included in the good prognosis group. Based on the miRNA expression data we were able to identify relevant predictors by applying variable selection methods (FCBF, Cox Univariate Regression, and ElasticNet). Subsequently, data were divided into training (70%) and test (30%) data sets. Nine machine learning algorithms (Support Vector Machines, Extreme Gradient Boosting Machine, Gradient Boosting Machine, Stochastic Gradient Boosting Machine, Random Forest, Conditional Random Forest, MLP (Multilayer perceptron), Generalized Linear Models, and Ranger Random Forest) were trained to build the prognosis predictor. Also, we identified the potential target genes for the miRNAs selected as relevant predictors through the analysis of differentially expressed genes. After applying the variable selection methods, 78 miRNAs were selected. Regarding the performance of the models on the test data set, the MLP, an artificial neural network model, presented the best metrics: 0.684 of AUC (Area Under the Curve); 0.740 specificity; 0.612 sensitivity; 0.603 of F1-Score and with an accuracy of around 70%. The miRNA-483, which was observed differentially expressed in more advanced stages of ovarian cancer in other studies, was considered to be an excellent predictor by the MLP model to classify patients in good or poor prognosis. When evaluating the top 1000 differentially expressed genes (655 up- and 345 down-regulated) in ovarian cancer samples compared with the healthy tissue, 196 were identified as potential targets for the 78 relevant miRNAs. Among them, 15 genes are regulated by the miRNA-483 and are associated with the prognosis and survival of patients with ovarian cancer.

Funding:

Link to Video: