Evaluation of lncRNAs as potential prognostic biomarkers in metastatic melanoma using machine learning

Cristiane Esteves Teixeira, Nicole de Miranda Scherer, Mariana Boroni, Guilherme Duarte de Moraes

Instituto Nacional de Câncer, INCA - Instituto Nacional de Câncer, Brazilian National Cancer Institute (INCA)

Abstract

Melanoma, despite having approximately 5% of incidence among skin neoplasms, is the most lethal of them. Although immunotherapies involving immune checkpoint blockade such as CTLA-4 and PD-1 have shown promise, there is a great variation in response to treatment, mainly related to genomic heterogeneity and immune infiltrates in patients. Therefore, prognostic biomarkers play a critical role in understanding disease progression and optimizing treatment and patient survival. Long non-coding RNAs (LncRNA) have already been described as prognostic biomarkers of several types of cancer, including cutaneous melanomas. In this sense, our work aims to investigate the role of lncRNAs as prognostic biomarkers specifically in melanoma patients who have metastasized, and to develop a model for predicting patient survival based on the expression profile of lncRNAs. We use machine learning (ML) techniques such as supervised machine learning algorithms on clinical-pathological data and gene expression of lncRNA from samples of patients diagnosed with metastatic melanoma obtained from the database TCGA (The Cancer Genome Atlas). A cohort of 348 patients was selected after filtering clinical data for metastatic samples only. Samples were divided into training and testing in a 4:1 ratio. We tested each of the 13, 954 lncRNA found in the samples using univariate cox regression to associate their level of expression with the overall survival of patients, resulting in 618 genes with $p < 0.05$. Ongoing analyses include the development of predictive models using three different regression algorithms (LASSO, Elastic Net and Ridge) in order to find a relevant lncRNA signature to classify patients according to the prognosis. The formula generated by the algorithms will be applied to calculate the risk of each patient and divide the cohort in low and high risk groups by the median. In addition, we will perform pathway enrichment analysis and generate a clinical nomogram.
Funding:
Link to Video: