Implementation of machine learning algorithms for medulloblastoma classification in a local Galaxy server

Luciane Sussuchi da Silva1, Leticia Ferro Leal1, Thais Talarico Hosokawa1, Murilo Machado1, Michael D Taylor2, Adriane Feijó Evangelista2, Rui Manuel Reis1

1Molecular Oncology Research Center, Barretos Cancer Hospital, Barretos, São Paulo, Brazil
2University of Toronto, Toronto, Ontario, Canada

Medulloblastoma Molecular Classification

Medulloblastoma is the most common malignant brain tumor in children, comprising around 20% of all childhood brain tumors. It has been classified in four main molecular subgroups with survival outcomes: WNT, SHH, Group 3 and Group 4 as summarized in Figure 1. Currently, there is no gold-standard method for diagnostic of these molecular subgroups, which could be efficiently applied in clinical practice.

22-gene Panel for PAM Algorithm

Northcott and colleagues** proposed a 22-gene panel for molecular classification of medulloblastoma from formalin-fixed paraffin embedded (FFPE) tissue blocks applying the NanoString nCounter® technology and the Prediction Analysis of Microarrays (PAM) algorithm. In their work, PAM was able to confidently predict around 88% of FFPE test samples using a training set of 101 and a validation set of 130 fresh-frozen samples. This pipeline has been applied at Barretos Cancer Hospital (BCH), Brazil. It’s being tested with a Canadian and a Brazilian training set of 240 and 130 fresh-frozen samples. This pipeline has been applied at Barretos Cancer Hospital (BCH), Brazil. It’s being tested with a Canadian and a Brazilian training set of 240 and 132 patients, respectively. Brazilian cohort representation is shown in Figure 2.

SVM and GNAS Expression

Support Vector Machine (SVM) method for medulloblastoma classification was developed for application in cases with inconclusive results for PAM algorithm. This tool is based in a wrapper for e1071 package and has been designed to work along with PAM tool in a clean interface. Moreover, additional genes like GNAS have been tested in order to improve prognosis prediction for SHH class and build a complete diagnostic tool. GNAS expression has been associated with overall survival rates in Brazilian cohort as shown in Figure 5 and may complement SHH prognosis.

Our Galaxy Experience

A Galaxy instance was recently implemented at Barretos Cancer Hospital, Brazil, and efficiently applied for research and diagnosis. Other research and diagnostic tools have been wrapped with Planemo and tested in a local ToolShed. Reproducibility, maintenance of history data and user friendly interface are key for gathering attention of both clinicians and researchers.