A Multi-signature Classifier Accurately Predicts Neuroblastoma Patients’ Outcome

Cornero A, Fardin P, Blengio F, Acquaviva A, Belli M L, Varesio L
Laboratory of Molecular Biology, Gaslini Institute, Genoa, Italy

Motivation
Neuroblastoma is the most common pediatric solid tumor of the sympathetic nervous system deriving from ganglionic lineage precursors. It is diagnosed during infancy and shows notable heterogeneity with regard to both histology and clinical behavior. Several gene expression-based approaches have been developed to stratify neuroblastoma patients. Most of these studies utilized gene signatures to improve the stratification of patients into risk groups and, although with mixed results, demonstrated a clear advantage of adding genomic analysis to risk assessment. Here we describe a new model to predict neuroblastoma patients’ outcome by taking into account a collection of published neuroblastoma related gene signatures.

Methods
Gene expression profiles of 182 neuroblastoma tumors were used to develop and validate an Artificial Neural Network classifier to predict patients’ outcome (alive or dead at 5 years after diagnosis). Patients were subdivided into three independent datasets including 60, 60, and 62 samples respectively. Forty-one gene signatures were selected from the literature by keywords search (‘neuroblastoma signature’ or ‘neuroblastoma expression profile’) and filtered according to availability of critical data. The procedure to generate the multi-signature classifier consists of two parts. First, an optimal classifier for each selected signature is generated. Second, the predictions obtained in the first part are combined to constitute the training and validation set for generating the multi-signature classifier. In details, for each signature a panel of twenty-three classification models were trained and tested on the first dataset and validated on the second dataset, based on the expression values of the genes included in each signature. The best performing classifier was selected for each signature. An accuracy cutoff of 80% was used to filter out the less informative signatures. The remaining classifiers, each corresponding to a single gene signature, were used in the second part of the analysis. Subsequently, a panel of twenty-three classification models were trained and tested on the results obtained in the previous step to predict patients’ outcome. Finally, these classifiers were validated on the third dataset and the model showing the best performance was selected.
Results
Twenty-four out of forty-one published gene signatures were capable to generate prognostic classifiers with an accuracy higher than 80% (ranging from 80% to 87%). Although these are very promising results, a classifier taking into account the different biological points of view carried by each of these signatures may produce a lower classification error. In fact, we show that an accuracy of 94% can be obtained by a classifier trained on the predictions generated by the signatures considered. These results demonstrate that a multi-signatures classifier can provide better results than what is possible to obtain from the use of each signature separately.

Contact email
andreacornero@ospedale-gaslini.ge.it