Multiclass Tumor Classification and Outcome Prediction of Central Nervous System Embryonal Tumors

Shayan Mukherjee
MIT, USA
sayan@mit.edu

Abstract

Two DNA microarray analysis projects are described.

In the first project: We asked whether the diagnosis of multiple common adult malignancies could be achieved purely by molecular classification, using DNA microarray-based tumor gene expression profiles. 218 tumor samples, spanning 14 common tumor types, and 90 normal tissue samples were subjected to oligonucleotide microarray gene expression analysis. The expression levels of 16,063 genes and expressed sequence tags were used to train and evaluate the accuracy of a multi-class classifier based on a Support Vector Machine algorithm. Overall classification accuracy was 78random classification. By contrast, poorly differentiated cancers resulted in low confidence predictions and could not be accurately classified according to their tissue of origin, indicating that they are molecularly distinct entities with dramatically different gene expression patterns compared to their well-differentiated counterparts.

In the second project: The categorization of embroyonal tumors of the central nervous system (CNS), a heterogenous set, is addressed and an attempt to predict patient's response to therapy for Medullobasltomas is made. These problems are addressed by developing a classification system based on DNA microarray gene expression data derived from 101 patient samples. We demonstrate that medulloblastomas are molecularly distinct from other brain tumors including primitive neuroectodermal tumors (PNET), atypical teratoid/rhabdoid tumors (AT/RT) and alignant gliomas. Previously unrecognized evidence supporting the derivation of medulloblastomas from cerebellar granule cells through activation of the Sonic Hedgehog (Shh) pathway was also revealed. We further show that the clinical outcome of children with medulloblastomas is highly predictable based on the gene expression profiles of their tumors at diagnosis.