Small Changes, Big Impact: miRNAs in Lung Cancer Detection

Oguz, Tarik (School: Del Norte High School)

MicroRNAs (miRNAs) with highly dysregulated expression levels, have so far constituted the focus of disease-specific biomarkers research. However, the complex regulating mechanisms between miRNAs and their target genes [4,6] along with the absence of a universally agreed upon framework to set fold change (FC) thresholds to conclude dysregulation, suggest miRNAs with mid-range dysregulation levels may yet carry clinically valuable information. Therefore, this study tests the hypothesis that 'miRNAs with lower dysregulation levels convey useful information for lung cancer (LC) diagnosis'. To this end, a dataset comprising full miRNA profiles for both a positive class (LC present) and a negative class (LC absent), is obtained. Statistically significant miRNA FC values are partitioned into three dysregulation ranges as follows: high (HD) FC<=0.5 or FC>=2.0, medium (MD) FC in (0.5,0.9] or in [1.1,2.0), and non-significant (ND) FC in (0.9,1.1). Using miRNA expression levels as features and incorporating a selection of the most discriminative 50 features per classifier, Random Forest binary classifiers for LC diagnosis are built, one for each of the three FC ranges, yielding the following classification accuracies: (HD) high 99%, (MD) low 99%, (ND) 92%. Furthermore, based on the pool of 100 miRNAs picked for HD and MD categories' classifiers, K-Means Clustering is used to analyze select sequence attributes of these miRNAs to understand how they relate to each other. The clustering model achieved the best performance (inertia knee-point based) with an n-value of 31, indicating cross-category sequential similarities as well as unique, novel MD clusters among selected miRNAs. (For cited references, please refer to the research plan.)

Awards Won: