Application of Biomedical Informatics to facilitate clinical use of gene expression microarrays in colon cancer

Guillermo López-Campos[a], Beatriz Pérez-Villamil[b], Alejandro Romera Lopez[b], Enrique Díaz Rubio[b], Fernando Martin-Sanchez[a]

[a] *Medical Bioinformatics Department, Institute of Health Carlos III, Majadahonda, Spain*
[b] *Molecular Oncology Laboratory, Hospital Clínico San Carlos, Madrid, Spain*

Abstract and Objective

Biomedical informatics plays a key role in extracting meaningful information and knowledge from microarray experiments that could be used in clinical settings. In this work we show an example of how biomedical informatics tools can be applied with the aim of facilitating the integration and analysis of different information levels in a microarray-based colon cancer gene expression study although there remain challenges to be addressed toward clinical use of gene expression microarrays. The main result of these analyses show that phenotypic characterization based on "Duke stages" does not match to different molecular entities.

Keywords:

Gene expression, Microarray, Biomedical informatics, Genomics.

Introduction

Traditionally, colon cancer classification (Dukes, AJCC) is based in clinical and morphological characteristics. Although this classification provides a good system that helps to take clinical decisions, is not enough to correctly predict patient clinical outcome. Therefore, Dukes A to D staging may not represent different molecular entities.

Biomedical Informatics (BMI) plays a key role in translational research bringing research results into clinical applications. In microarray analysis BMI this role is essential to overcome some of the pitfalls related with the use in clinics of microarrays.

In our study, microarray analysis has been carried out in colon tumor samples with the aim to obtain homogeneous groups in colon tumors, based on their expression patterns, and to analyze whether Dukes stages correlate with the microarray molecular classification.

Materials and Methods

A total 88 Agilent whole genome microarrays were used to analyze gene expression in 88 colon tumors (tumor-Cy5/pool-Cy3). Microarray data were normalized using lineal-LOWESS and filtered to remove those genes with low fluorescence or those with a flat pattern.

An in house database was developed for the clinical annotation of the analyzed samples.

Unsupervised analysis was carried out applying hierarchical clustering techniques centering the genes and using a centered Pearson correlation as distance metrics and average-linkage. Lastly, a supervised analysis for class comparison controlling the False Discovery Rate between the 4 defined Dukes stages and for the novel groups identified by the class discovery results was completed, along with these comparisons pathways functional analysis was done.

Results

Data filtering allowed us to reduce the number of genes to 1722. Using those genes class comparison between the four Dukes stages showed no significant differences ($p=0.377$) between A, B, C or D classes.

Hierarchical clustering for molecular classification of the 88 colon cancer samples allowed us to identify 4 clusters. Dukes stages A, B, C, D showed a disperse pattern and did not coalesce in any of the clusters. These new clusters were statistically robust (Robust index >0.75) and biologically meaningful.

Using pathway analysis we have identified 20 KEGG pathways affected among the new identified clusters.

Conclusion

Our data suggest that Dukes A to D stages are not different molecular entities and may explain why they display a diverse clinical behavior or respond differently to treatment.

Biomedical informatics is crucial to facilitate an integrative multilevel data analysis, achieving semantic integration of heterogeneous data sources as well as for managing expression profile data as sub-phenotypes – improving clinical annotation of samples and supporting association studies.