CALIFORNIA STATE SCIENCE FAIR
2012 PROJECT SUMMARY

Name(s)

Natalie Ng

Project Number

S0519

Project Title

Development of a Novel Biomarker Discovery Tool to Identify Clinical Signatures from Deconvoluted Expressions

Abstract

Objectives/Goals

Tissue heterogeneity is a major confounding factor in microarray-based gene expression analysis. Gene expression deconvolution is an innovative method to overcome this problem by decomposing the global gene expression into pure cell expression subprofiles. Differential analysis can then be performed on the deconvoluted expressions to identify disease-related genes that may otherwise be undetectable. The primary objectives of this project are the development of a novel biomarker discovery tool to be used in conjunction with statistical gene expression deconvolution to identify prognostic signatures for breast cancer patients. This project has two major components: (1) development of a novel biomarker discovery tool and, (2) development and application of a statistical deconvolution method to identify gene signatures which can be used as prognostic predictors of disease outcomes.

Methods/Materials

Microarrays of Affymetrix Human Genome U133A Array platform were used in the cell type biomarker discovery workflow. Cell type specific biomarkers were identified using statistical significance tests (ANOVA and TukeyHSD) and up-regulation ratio computation. Microarray dataset GSE2034, from a published breast cancer study, was used to demonstrate the application of gene expression deconvolution to identify clinical signatures. An automated statistical gene expression deconvolution procedure was developed based on a two-step iterative algorithm implemented in Matlab.

Results

Biomarkers of immune, stromal, and tumor cell types have been identified and verified. Correlations of cell type biomarkers were excellent in the training and validation sets. The biomarkers provided biological identification of the deconvoluted expressions to their corresponding cell types. Using statistical gene expression deconvolution, gene signatures that can discriminate breast cancer patients according to clinical outcomes (relapse versus relapse-free) have been identified.

Conclusions/Discussion

A novel biomarker discovery tool and an automated statistical gene expression deconvolution procedure were developed to analyze global gene expressions of breast tumor samples to identify prognostic gene signatures predictive of clinical outcomes. Functional analysis of prognostic signatures provided insights into molecular pathways associated with tumor progression and metastasis.

Summary Statement

I designed and developed a novel biomarker discovery tool and an automated statistical gene expression deconvolution protocol to analyze breast tumor samples and identify prognostic gene signatures predictive of clinical outcomes.

Help Received

My mentor, Dr. Neta Zuckerman (Stanford University, City of Hope Hospital) provided guidance and feedback. Professor Peter Lee (Stanford University, City of Hope Hospital) provided a summer internship position.