Poster L-23

Novel Gene Expression Analysis Using SAM and PAM Algorithms to Determine Estrogen-regulated Genes Involved in Response to Neoadjuvant Endocrine Therapy for Breast Cancer

Authors:
Susan Trapp (*"Computational Bioscience Program, Univ. of Colorado Health Sciences"*)
Tzu Phang (*"School of Medicine, Univ. of Colorado Health Sciences"*)
Djuana Harvell (*"Division of Endocrinology, Univ. of Colorado Health Sciences"*)
Larry Hunter (*"Computational Bioscience Program, Univ. of Colorado Health Sciences"*)
Anthony Elias (*"Dept of Oncology, Univ. of Colorado Health Sciences"*)
Jennifer Richer (*"Division of Endocrinology, Univ. of Colorado Health Sciences"*)

Short Abstract: We present a gene expression analysis approach incorporating SAM and PAM algorithms to determine estrogen-regulated genes involved in response to hormonal therapy for breast cancer. The challenge is can we accurately predict gene biomarkers based upon drug treatments and response from a dataset of 7 paired breast cancer patient samples (pre- and post-treatment).

Long Abstract:
Motivation: Preoperative (neoadjuvant) hormone therapy in patients with locally advanced, estrogen receptor (ER) positive breast cancers is known to increase the rate of successful breast conserving surgery in patients without adversely affecting event-free and overall survival. Aromatase inhibitors and tamoxifen (Tam) both target the ER pathway and have been shown to be individually effective for preoperative use. The presence of ER or progesterone receptors (PR) is a known predictor of responsiveness and approximately 70% of tumors that have ER or PR do regress in response to these preoperative hormone therapies. However, at least 30% of receptor positive tumors do not respond to hormone treatments. We postulate that specific genes in the expression profile of breast cancers can better predict response to preoperative endocrine therapy and also that genes and their protein products that change with successful shrinkage of tumors may themselves be candidates as direct targets for future therapies. With the exception of one other study, previously reported studies have analyzed microstatic tumors; this study takes advantage of the dynamic profile of each patient's tumor before and after treatment. Although we have fewer samples (7 patients) this study has the potential to be much more informative due to the power of paired samples from the same patient.
Methods: To begin to identify such genes, paired pre- and post-treatment tumor specimens were obtained from patients with locally advanced, ER+ and/or PR+ breast cancers enrolled in a clinical trial evaluating the effectiveness of a steroidal aromatase inhibitor (exemestane) with or without tamoxifen. The tumors were analyzed by expression profiling using an Affymetrix U133Plus2 microarray platform (54,647 probe sets). Previous Results: Unsupervised cluster analysis resulted in pre- and post-treatment biopsies from individual patients clustered together indicating that each individual's tumor samples most closely resemble each other. A preliminary Students t-test analyses identified 846 genes

differentially expressed genes changing in pre- vs. post-treatment samples ?1.5 fold in 7 of 7 ($p < 0.05$), indicating that ~ 1.5% of the total genes assayed change with treatment. Supervised hierarchal clustering with this subset of E-regulated genes associated with tumor shrinkage, sorted patient tumors samples into pre-versus post-treatment groups and further segregated tumors that responded well to treatment from those that responded poorly. Current Results: A bioinformatics approach was performed utilizing SAM (Significant Analysis of Microarray) and PAM (Prediction of Analysis of Microarray to 1) confirm which genes are changing with treatment [exemestane with or without Tam], and 2) determine if it is possible to predict, based solely on data from pretreatment biopsies which tumors will shrink in response to treatment versus those that will not shrink or that will progress. SAM assigns a score to each gene on the basis of change in gene expression relative to the standard deviation of repeated measurements. For genes with scores greater than an adjustable threshold SAM uses permutations of the repeated measurements to estimate the percentage of genes identified by chance utilizing FDR (False Discovery Rate). PAM utilizes the nearest shrunken centroid methodology to carry out sample classification from gene expression data. For our first aim, 51 genes with FDR of 22% were identified. Many of these genes are known to be regulated by ER (which is a transcription factor). Further biological investigations are underway to follow up on interesting genes with unknown function. Following the PAM analysis on the 51 genes selecting a threshold of 1.8 identified 23 putative genes as potential markers of treatment response. PAM analysis to predict response, aim two, is currently being investigated. Conclusions. This project has potential to help clinicians determine which women are good candidates for hormonal therapy. If genes identified in aims 1 and 2 are confirmed to be experimentally significant in an independent data set, these will be potential new therapeutic targets and better predictors of response to hormonal therapy. In the future, we will investigate multivariate methods, such as 2-way ANOVA, to study the dependency between different factors (treatment versus response). The current classification method (PAM) is limited to univariate comparisons (pre- versus post-treatment); we will examine ways to take advantage of the multivariate nature of the dataset.