

Discovering breast cancer prognostic biomarkers using a novel feature selection tool

Jie Cheng
Quantitative Sciences
GlaxoSmithKline
Collegeville, PA, USA
jie.j.cheng@gsk.com

Abstract — We will present a case study of applying a novel feature selection tool to breast cancer biomarker discovery. Using a publicly available gene expression microarray dataset, we discovered prognostic biomarkers for various patient subpopulations stratified by clinical variables. We then used independent datasets consist of lymph node negative patients to validate 20 potential biomarkers. The results show that our 20-gene signature as well as many of the discovery individual prognostic biomarkers can achieve comparable or better performance compared to the clinical or gene signature based prognostic risk scores, especially for young ER+ patients. These discovered biomarkers have the potential to be used in clinical settings to identify a subset of the lymph-node-negative (Node-) and estrogen-receptor-positive (ER+) patients who are at a higher risk of relapse and should be treated more aggressively. We will also discuss good practices in industrial biomarker discovery. (*Abstract*)