Using copulas to select prognostic genes in melanoma patients

Linda Chaba, John Odhiambo and Bernard Omolo

Abstract

Melanoma of the skin is the 5th and seventh most commonly diagnosed carcinoma in men and women, respectively, in the USA. So far, gene signatures prognostic for overall and distant metastasis-free survival, for example, have been promising in the identification of therapeutic targets for primary and metastatic melanoma. But most of these gene signatures have been selected using statistics that depend entirely on the parametric distributions of the data (e.g. t -statistics). In this study, we assessed the impact of relaxing the parametric assumptions on the power of the models used for gene selection. We developed a semi-parametric model for feature selection that does not depend on the distributions of the covariates. This copula based model only assumed that the marginal distributions of the covariates are continuous. Simulations indicated that the copula-based model had reasonable power at various levels of the false discovery rate (FDR). These results were validated in a publicly available melanoma dataset. Relaxing parametric assumptions on microarray data may yield procedures that have good power for differential gene expression analysis.