Modeling Distant Metastasis-Free Survival: Applications to Hazard Prediction and Pairwise Gene Interaction Discovery

Yang, Russell (School: The Harker School)

Models of metastasis-free survival based on microarray gene expression can be powerful tools for clinicians and have broad applications in risk quantification, genetic research, and palliative care. This project focused on developing various survival models for distant metastasis-free survival and exploring some of the applications of those models in genetic and clinical settings. Several linear and nonlinear dimensionality reduction methods were employed to transform a set of 20,000+ genes to a smaller dataset. Various survival models were considered, including the traditional semiparametric Cox Proportional Hazards Model, nonparametric ensemble-based methods, and a Cox Proportional Hazards Deep Neural Network. Each combination of dimensionality reduction and survival model was evaluated using a nonparametric bootstrapping approach (B=100). Concordance indices were used to assess model prediction ability. The assumptions of the Cox Proportional Hazards Model were rigorously tested using Schoenfeld, Deviance, and Martingale residuals. Next, an algorithm was devised to identify potential pairwise epistatic gene interactions between MYC (a proto-oncogene) and other genes in the dataset. A Kolmogorov-Smirnov test was performed, showing a highly statistically significant difference in gene z-score density upon the addition of MYC to the model. Lastly, preranked gene set enrichment analysis was performed on significant model-identified genes (with Hallmark and C1 positional sets as references) to find pathways and their associated normalized enrichment scores. The models developed in this study can be used to estimate the prognostic effect for any gene of interest. Perhaps more importantly, the models can also be used for survival prediction in a palliative setting.

Awards Won: