**A Novel Bioinformatics Approach to Estimate Hazard in Pediatric High Grade Glioma**

Gregory Schwing, Marc Moisi, Chindo Hicks

Affiliations  
Gregory Schwing, Wayne State University School of Medicine, Detroit, MI;

Marc Moisi, Wayne State University School of Medicine, Detroit, MI;

Chindo Hicks, Louisiana State University Health Sciences Center – New Orleans, New Orleans, Louisiana

Corresponding author  
Gregory Schwing, BS, BS, Wayne State University School of Medicine, Detroit, MI, 48201  
telephone: 504-994-0403; fax: xxx-xxx-xxxx; e-mail: gregory.schwing@med.wayne.edu

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Introduction: “Cox-nnet” is an Artificial Neural Network that performs Cox (Proportional Hazards) Regression. The neural network learns through training, optimizing, and testing to find hidden connections between variables to estimate hazard.

Methods: The combined mRNA Expression z-Scores from Affymetrix arrays, Agilent arrays, and RNASeq of Mackay et al. 2018 were retrieved from pedcbioportal(dot)org. The inclusion criteria were the reporting of clinical covariates – Diagnosis, WHO Grade, Tumor location, Gender, Age, Censored Survival, and Histone subtype. The Samples meeting inclusion criteria (n=162) were analyzed with Cox-nnet, using the expression z-scores of 8,540 genes and the clinical covariates as inputs.

Results: The Kaplan-Meier curves generated by partitioning the test set above and below estimated median log hazard ratio (MLHR = 0.04) were statistically significant by log-rank test (Chisq= 7.4 on 1 degrees of freedom, p= 0.006). The five most important clinical covariates (weight given in parentheses) were WHO Grade (0.078), Histone 3.3 K27M mutation (0.025), Wild-Type Histone (0.0156), Hemispheric (0.009), and Midline (0.008). The five most important genes were those encoding for Perforin-1 (0.674), Neuropeptide-Y (0.498), DRC11 Antibody (0.280), Serine Protease 3 (0.280), and Zinc-Alpha-2-Glycoprotein-1 (0.222). The three most important pathways were Hepatic Fibrosis (2.021); Role of Macrophages, Fibroblasts and Endothelial Cells in Rheumatoid Arthritis (1.848); and Axonal Guidance Signaling (1.617). The pHGG-specific pathways were Glioblastoma Multiforme Signaling (0.785), Glioma Invasiveness Signaling (0.449), and Embryonic Stem Cell Pluripotency (0.630).

Conclusions: The predictive model was strong, as indicated by the significant log-rank test. Of the top five genes, four are either immunologic or CNS-related. Only one, Zinc-Alpha-2-Glycoprotein-1, is tumorigenic. The pathways follow a similar trend. This partially explains the difficulty in treating pHGG. Most of the important features are either immunologic, a field in the early stages of drug development, or related to the underlying CNS physiology, which could be devastating to disrupt.