**shinyGEO: a web application for analyzing Gene Expression Omnibus (GEO) datasets using shiny**

**Jasmine Dumas1\*, Michael Gargano2, Garrett M. Dancik2**

1. College of Computing and Digital Media, DePaul University, Chicago, IL

2. Department of Mathematics and Computer Science, Eastern Connecticut State University, Willimantic, CT

\*Contact author: jasmine.dumas@gmail.com

**Keywords:** Bioinformatics, Gene Expression, Survival Analysis, Shiny

Identifying associations between patient gene expression profiles and clinical data provides insight into the biological processes associated with health and disease. The Gene Expression Omnibus (GEO) is a public repository of gene expression and sequence-based datasets, and currently includes >42,000 datasets with gene expression profiles obtained by microarray. Although GEO has its own analysis tool (GEO2R) for identifying differentially expressed genes, the tool is not designed for advanced data analysis and does not generate publication-ready graphics. In this work, we describe a web-based, easy-to-use tool for biomarker analysis in GEO datasets, called *shinyGEO*.

*shinyGEO* is a web-based tool that provides a graphical user interface for users without *R* programming experience to quickly analyze GEO datasets. The tool is developed using '**shiny'**, a web application framework for *R*. Specifically, *shinyGEO* allows a user to download the expression and clinical data from a GEO dataset, to modify the dataset correcting for spelling and misaligned data frame columns, to select a gene of interest, and to perform a survival or differential expression analysis using the available data. The tool uses the Bioconductor package '**GEOquery'** to retrieve the GEO dataset, while survival and differential expression analyses are carried out using the '**survival'** and '**stats**' packages, respectively. For both analyses, *shinyGEO* produces publication-ready graphics using '**ggplot2'** and generates the corresponding *R* code to ensure that all analyses are reproducible. We demonstrate the capabilities of the tool by using *shinyGEO* to identify diagnostic and prognostic biomarkers in cancer.