



Gene Set Enrichment Analysis in Pediatric Cancer with Multivariate Logistic Regression
Lucy Wu (EAS 2018)
Advisor: Deanne Taylor

I had the pleasure of working under Dr. Deanne Taylor of the Children's Hospital of Philadelphia this summer. The goal of my project was the development of computer code for conducting more accurate and efficient analyses of neuroblastoma gene sets using R. The code will eventually be implemented in an online web tool for gene set analysis.

An existing method called Gene Set Enrichment Analysis (GSEA) summarizes gene-level statistics at the level of predefined gene sets, such as signaling pathways. Essentially, a scientist uses GSEA, for a given gene set, to compare whether and how accurately the genes he discovered to be in that gene set match up with the genes in the predefined gene set. If there is any difference, then the scientist will know it can be attributed to different variables, or phenotypic expressions, of the scientist's experimental genes, and GSEA will analyze the degree of difference. However, there are two main limits to GSEA. First, GSEA only takes into account the effect of a single variable, and possible confounding variables that bring bias into the statistics are not considered. Second, existing applications for statistical computation contain extraneous code that would not be applicable to this particular study and run in a linear stepwise manner, thus adding unnecessary system run time.

Therefore, my project focuses on 1) the mvGSEA (multivariate) method by using logistic regression models that take into account effects of multiple and confounding variables, and 2) reducing existing regression code and developing additional code for efficient system run time through parallelization, as opposed to the current stepwise approach.

Through my research experience, I learned more about statistical logistic regression and the code behind the computation. I developed insight on how to write code more efficiently so as to minimize space and time. I also gained more patience in checking code more carefully, and learned the importance of asking for help when needed.

Participating in this project has made me more eager for a future in science and investigation as a pediatrician. It has made me become more motivated to tackle seemingly convoluted problems, which will help me as a student and in my career. I would wholeheartedly like to thank everyone who made this opportunity possible for me.