FlyRNAi Enrichment Tool: Tissue Preferred Expression Analysis for Drosophila

Chao, Tiffany (School: Blind Brook High School)

Gene Enrichment Analysis is used to analyze the relationship between genes and known biological annotations. The lack of a versatile Gene Enrichment Tool has been a major hurdle throughout bioinformatics studies on Drosophila melanogaster. This project is part of an initiative to devise a novel gene enrichment analysis tool tailored to Drosophila, which will enable scientists to quickly screen and identify candidates potentially involved in specific tissues, pathways, biological processes, etc. In this study, 29 Bulk RNA-seq datasets, generated from various dissected tissues, were analyzed. This study used Python to design and implement a data mining application used to group data records for tissues based on annotation file. An application of this nature was previously unavailable under public domain. Statistics including mean, standard deviation, and p-value were calculated for each group and group set. The application looped through 17,742 genes in the datasets, compared the statistics, and identified genes with an expression level 3x higher in one tissue than in any other tissue groups. The application identified 3764 genes in the output. The results were validated by querying a subset of the identified genes in BioLitMine, an advanced literature mining tool, and Flybase, a Drosophila genetics database. High correlation between these sources confirms the accuracy of this application. The application was published on Github, which may be useful in future studies involving data analysis processes. Expression data for these tissue-preferred genes were stored in the FlyRNAi database, and were used to construct the tissue-specific enrichment analysis function in the new Gene Set Enrichment Tool.

Awards Won: