Investigating Cancer Mutations: Improving the Analysis of Cancer Data with Software

Cunningham, Caroline (School: Washington-Lee High School)

DNA sequencing of patient's tumors allows computational methods to reveal unprecedented insights on their genomes. Analysis of mutations in tumor genomes allows inference of the underlying mutational processes. Knowledge about mutational processes is lacking, and current computational methods are limited in scope of analysis and do not have clearly-defined methods. This project implements a comprehensive, hypothesis-driven approach to mutational pattern analysis through the development of a cancer analysis framework that analyzes the presence of 5 mutational patterns, APOBEC, UV, Tobacco, AID, and Aging, in cancer mutational data. The program tests the hypothesis that a given mutational pattern is significantly associated with the mutational processes in tumor samples and reports enrichment, mutational burden, and a Chi-squared p-value, thus quantifying mutational pattern prevalence. From analyses of over 5000 cancer genomes from 15 cancer types, the created framework identified 39 significant associations between the cancers and the 5 known mutational patterns. These findings corroborated several previously-characterized associations, which act as a validity test for the analysis framework. New associations were also identified, particularly for ten cancer types that were found to be associated with more than one mutational pattern, suggesting the etiologies are more complex than one origin. Thus, the mutational pattern analysis tool strengthens the understanding of the mutational processes occurring in a cancer sample and establishes an improved framework to analyze mutational data. Moreover, the discovered associations have research and clinical implications. The software will be integrated into a bioinformatics software package, easily accessible over the internet.

Awards Won: