Title

Examining genomic data to understand functional effects of genetic variants in ovarian cancer

Authors

Omodolapo Nurudeen,[1] Amanda Bataycan,[2] and Ming-Ying Leung[1,2,3]

[1]Bioinformatics Program, [2] Computational Science Program, and [3]Department of Mathematical Sciences, The University of Texas at El Paso, El Paso, TX

Abstract

Cancer is a global health priority, with ovarian tumors being the fifth in cancer deaths among women. Ovarian cancer poses serious dangers to patients' well-being, with the year relative survival rate at 50.8.%. The goal of this research is to identify yet unreported possible ovarian cancer-related single nucleotide variants (SNVs) for downstream bioinformatics and experimental investigations to gain better understanding of the molecular underpinnings of this disease. We developed custom Python codes to convert variant call format (VCF) files obtained from the Genomic Data Commons Portal into more readable CSV files that provided critical information of the SNVs, including their genomic locations and associated transcript IDs as well as the reference and mutated bases. After processing and cleaning the CSV files from 426 patient records, we compiled 213,894 distinct variants that occur in tumor samples only, 78 in normal samples only, and 8,858 common to both. Next, we used the FATHMM-XF web server, which is based on a supervised machine learning approach, to score the SNVs on a scale of 0 to 1, with 1 being most pathogenic. Out of all the SNVs, 52,802 were found within coding regions, and 17,909 of them were classified as pathogenic with a score of above 0.5. In the near future, we will incorporate other functional effect assessment tools such as SNPnexus and PROVEAN with FATHMM-XF to establish a scoring scheme for evaluating the cumulative impacts of SNVs on genes, with the ultimate aim of revealing novel gene targets for cancer therapy.