Title

Scoring Single Nucleotide Variant to Predict Novel Thyroid Cancer Genes

Authors

Omodolapo Nurudeen[1], Amanda Bataycan[2], and Ming-Ying Leung[1,2,3,4]

[1]Bioinformatics Program, [2]Computational Science Program, [3]Department of Mathematical Sciences, [4]Border Biomedical Research Center, The University of Texas at El Paso, El Paso, TX

Abstract

It is estimated that over 50,000 Americans are newly diagnosed with thyroid cancer (THCA) annually. Although treatments for THCA are mostly successful, about 2,000 people die from the disease every year. The goal of this research is to identify novel genes associated with THCA for further investigation. Through the Genomic Data Commons (GDC) Portal, we obtained 504 data files in variant call format (VCF) containing single nucleotide variants (SNV) information from 477 patients with THCA. We quantified the deleterious impact of each SNV using four functional effect analyzers: FATHMM, SIFT, CADD, and POLYPHEN. As these software tools were developed based on different algorithms, their assessments of deleterious effects of the SNVs vary substantially. We have developed a cumulative scoring function, called Q(Gene), that calculates the overall pathogenic impact caused by the nonsynonymous SNVs on each protein-coding gene based on the different functional effect assessments while considering the SNVs' occurrence frequencies in tumor and normal samples and transcript lengths of the genes. The Q(Gene) scoring resulted in a ranked list of 6408 possibly pathogenic protein-coding genes. Cross-referencing THCA related genes from public databases and published literature, we found 51 out of the top 1% gene to be novel. and conducted additional bioinformatics analyses on them using GO term enrichments. Some of the top GO terms represented by the novel genes include cytoskeleton organization, actin filament-based activities, calcium ion transmembrane import, and muscle tissue development. We plan to use this gene list and GO terms for future in-silico and wet-lab investigations.