Using a Sequence-to-Structure-to-Function Approach to Functionally Characterize Protein Coding Missense Mutations in the Human and Rat Genomes

The Medical College of Wisconsin

PI: Jeremy W. Prokop

Grant Number: 1-K01S25435-01

The goal of this BD2K K01 application is to provide Dr. Jeremy W. Prokop with the mentoring and training to be an independent investigator in big data to knowledge. Mentoring and training will be provided to Dr. Prokop in genomics, proteomics, computer science, and statistics to advance his skill to allow for independence. This will allow for further develop of his sequence-to-structure-to-function analysis for interpretation of protein coding genetic variants into a usable workflow available to other scientists. The mentoring throughout this award will be from Dr. Howard Jacob, a leader in rat/human genetics and tool development to understand variants in whole genomes. Additionally, Dr. Andrew Greene serves as a co-mentor to help advance skills in proteomics and Dr. Christina Kendziorski as a co-mentor to advance statistical tools/approaches. The Medical College of Wisconsin (MCW) is a leader in the use of whole genome sequencing in understanding human health, with a Cap/CLIA certified clinical sequencing facility, tools for identifying genetic variants from whole genomes (such as Carpe Novo), and operation of databases like the Rat Genome Database. These tools and knowledge at MCW will serve as an asset for the training and completion of the Aims in this award. Dr. Prokop's research focus in this grant is to further expand the sequence-to-structure-to-function approaches he developed during his Ph.D. and initial postdoc into a workflow and web submission server for other users. Aim 1 of the grant organizes these steps into a workflow allowing for the development of the web based submission server. To test the workflow, 75 candidate genes for cardiovascular disease will be screened. Initial use of the approach has revealed hypotheses for how genetic variants in Havcr1 and Shroom3 result in altered cardiovascular drug response or disease. Variants in these two proteins will be biochemically characterized using a novel decision tree to standardize experiments (Aim 2) and to serve as a quality control for the approaches of Aim 1. From the results of the 75 screened cardiovascular genes, the four genes with the highest confidence score will be validated using the biochemical decision tree in year four and five of this award serving as an additional quality control for the approaches in Aim 1. This grant will provide the training and support for Dr. Prokop to be integrated into the Medical College of Wisconsin's Clinical Sequencing program, allowing for additional R01 proposals for validation of genetic causes of disease, and facilitate the development of Dr. Prokop into an independent researcher in the use of big data. PUBLIC HEALTH RELEVANCE: Development of a new tool package that is able to interpret genetic variants of disease genomes into a testable hypothesis of how protein function is perturbed. This tool will allow for improved diagnostic and treatment methods for many human diseases that employ whole genome/exome sequencing such as cardiovascular, cancer, and rare diseases.