STRONG CHILDREN'S RESEARCH CENTER

Summer 2015 Research Scholar

Name: Francis (Frankie) James
School: Loyola Marymount University
Mentor: Alex R. Paciorkowski, MD

ABSTRACT

Title:
SOLVE-Brain: Identification of variations in novel genes for developmental brain disorders

Background: Whole exome sequencing (WES), has identified pathogenic variations in genes for developmental brain disorders such as epilepsy, autism, and intellectual disability. Downstream analysis of WES data includes pedigree filtering and annotation and remains a challenge. SOLVE-Brain is a bioinformatics toolkit designed to simplify aspects of downstream analysis of WES data for developmental brain disorders.

Objectives:

- Improve efficiency and ease of use of SOLVE-Brain
- Test these improvements in a new cohort of WES data from individuals with developmental brain disorders

Methods: SOLVE-Brain is a suite of tools that allow pedigree filtering as well as annotation using a variety of user specified inputs. Several improvements were introduced to allow more flexible and accurate pedigree filtering. Functionality was added to allow the user to identify variants unique to a proband of interest within a cohort. Additionally, output format was changed to allow easier visualization of candidate variants.

To test these improvements, we then analyzed 21 WES datasets from individuals with infantile spasms and 14 WES from individuals with developmental brain disorders. We also reanalyzed two prior unsolved WES datasets with infantile spasms. For WES, paired end 100 bp reads were obtained on an Illumina sequencer and analyzed with BWA, Picard, GATK, and Annovar. Analysis was performed on BlueHive, a 178-node Linux cluster at the University of Rochester's Center for Integrated Research Computing. Candidate variants identified were confirmed using standard Sanger methods.

Results: Overall WES analysis, from mapping to variant calling on average required ~96 hours in the BlueHive environment. After that, downstream analysis (annotation, pedigree filtering, and candidate variant identification) required a further 48 hours. The improved functionality added to SOLVE-Brain allowed an estimated decrease of at least 1 hour per WES data set. For this group of data, this translates into a reduction of about 37 hours of work. In addition, the improved output format adds important qualitative ease-of-use improvements.

Our analysis identified strong candidate variants in *CPZ, DYNC1L1, OSCP1, RABGAP1, OBSCN, PEMT,* and *PAPLN*. In the developmental brain disorder group, we identified a strong candidate variant in *CLCN6*. And in the reanalysis of the previously unsolved infantile spasms data sets, a new X-linked variant in *SRPK3* was identified in one family, and new autosomal recessive variants in *PAPLN* that are currently undergoing confirmation.

Conclusion: WES is a valuable tool for the identification of novel causative genes in human disorders, including autism, epilepsy, and intellectual disability. Due to the relatively large volume of data WES produces, optimizations in data analysis will improve the speed that diagnoses can be made. We have made optimizations to SOLVE-Brain that decreased downstream analysis by 77%. This allowed for faster progression from mapping to final candidate variant identification, and therefore diagnosis. In addition, these improvements allowed the identification of new candidate variants in WES data sets that were unsolved by previous versions of the software. This allows for more rapid genetic diagnosis in a class of severe disorders of the developing brain.