

Analysis of short-read data from NCBI SRA and CGHub/TCGA using HIVE-XLD: Evaluation of the effects of non-synonymous variation in the Human proteome

Raja Mazumder¹ and Vahan Simonyan²

¹George Washington University, USA

²Food and Drug Administration, USA

Current sequencing technologies are generating petabytes of data that are inaccessible to majority of the research community because of the costs and expertise required to analyze big data. Another roadblock to analyzing such data is the lack of curated information in NGS data repositories such as NCBI SRA. To address the above challenges, we have implemented a low-cost High-performance Integrated Virtual Environment of eXtra-Large Data (HIVE-XLD) private cloud at GWU and US FDA. Effects of variation on active sites and glycosylation sites will be presented to illustrate the power of integration of big-data with functional objects such as active sites, binding sites and pathways.

Biography

Raja Mazumder has applied bioinformatics and computational biochemistry strongly rooted in evolutionary biology form in his research program. His current research goals includes development of methods to perform analysis of extra-large data sets (such as next generation data) within the context of evolutionary systems biology to identify experimental therapeutic targets, and to create a bioinformatics data-warehouse for "omics" data integration. He is currently involved in implementing High-performance Integrated Virtual Environment (HIVE) (<http://hive.biochemistry.gwu.edu>), a cloud-based environment that contains various bio-scientific tools for analysis of extra-large data; identification of vaccine and therapeutic targets for Hepatitis C Virus; identification of non-synonymous Single Nucleotide Variations (nsSNV) that affect post-translational modifications.

mazumder@gwu.edu