

Statement of work

The Gerstein Lab will develop a framework to functionally prioritize SVs over five contexts: (1) Impacting protein coding genes; (2) Impacting non-coding RNAs; (3) Impacting non-coding regulatory regions such as Transcription Factor Binding Sites (TFBS); (4) Impacting regions associated with eQTLs and; (5) Impacting regions associated with allele activity. The impact score will take into account the varied ways a SV can affect genomic elements (e.g. partial overlap v engulf) and will integrate conservation information, existing genomic annotations, epigenetic and transcriptomic datasets from sources such as ENCODE, 1000 Genomes, GTEx and, RNA-seq data generated by TopMed. Furthermore, we will upweight the impact score of SVs overlapping elements with ubiquitous activity, high network connectivity (i.e. hubs) and strong allelic activity (i.e. demonstrated functional sensitivity to variants).