**Uncovering Predisposing Variants in The Cancer Genome Atlas (TCGA)**

The Cancer Genome Atlas PanCanAtlas Germline Working Group

Sequencing of normal, non-neoplastic samples from individuals with cancer in The Cancer Genome Atlas (TCGA) can be used to discover heritable factors in cancer susceptibility. The TCGA PanCanAtlas Germline Working Group (GWG) characterized germline variants in 14,403 samples across 32 cancer types. We dockerized the GenomeVIP variant calling tool (github.com/ding-lab/GenomeVIP/) and deployed over 40,000 virtual machines on the ISB Cancer Genomics Cloud (ISB-CGC) during the course of the project, amounting to ~332,000 wall-hours of cpu time to generate ~5.9 million variant calls, which will be further filtered and annotated for downstream analysis. The coordination of GWG exemplifies a much needed sustainable model of cloud-enabled data sharing and­­­­­ computation as genomic projects expand exponentially in scale.

Based on the shared variant calls, we are conducting burden and allelic imbalance tests to uncover predisposition genes. Further, we will examine the relations between germline and somatic mutations at both the sample and cohort levels. Critically, we will associate known germline pre-disposition variants from GWAS studies and rare variants in DNA Damage Response with specific somatic mutational phenotypes. In addition, we will investigate germline miRNA and regulatory variants across all samples by combining analysis with epigenome data from ENCODE. Analysis of clinical data show a subset (~658) of cases in TCGA have at least 1 first-degree relative with cancer. We plan identify familial relations within the dataset using array data and uncover potentially segregating variants.

The TCGA PanCanAtlas Germline Working Group dataset and analysis will provide the largest-to-date resource for understanding the contribution of germline variants to cancer. The GWG will also create a cloud-based web portal for investigators to independently explore the germline calls and their associations with molecular and clinical data in TCGA.