Project Proposal – Yang’s Gang

General: Pipeline for very-low frequency variant calling of somatic mutations

In general, somatic variant callers have been directed at detection of non-reference alleles (NRAs) present in tumor from surgically acquired tissue. Currently, the common practice is to limit variant detection to allele frequency >5% although some more recent applications have used >1% as a threshold. Thus, commonly used variant callers (e.g., Strelka2, Mutect2, LoFreq) for ‘low’ frequency variant detection do not interrogate NRAs with a variant allele frequency (VAF) < 1-3%. The development of a variant caller pipeline to adjudicate NRAs with a VAF < 1% based on pathogenicity of the variant rather than VAF has applications in understanding tumor phylogeny and the overall mutational landscape in solid tumor DNA through improved sensitivity. In addition, a pipeline to detect very-low frequency variants would have applications in tumor-derived cell-free DNA from plasma where VAF is commonly <0.5%, particularly in cell-free DNA associated with non-metastatic solid tumors.

Here, we propose to develop a pipeline using a python wrap around each module to handle the input/output for each module that uses java-based software to perform each computational or analytic step. To evaluate performance, sensitivity will be measured using a BAM files of cell-free DNA (coverage: ~3000X) from a healthy control where COSMIC exonic variants were synthetically injected across a range of VAFs. A .vcf of the injected positions and associated variants will be used as reference to determine sensitivity. Specificity will be measured using a BAM file from a separate healthy control sequenced under identical conditions. For all data sets, germline DNA is available as a reference to identify SNPs. The following workflow is proposed:

BAM --> GATK\* (variant caller) --> Isolation of exonic positions --> Snpeff\*\* (variant effect prediction) --> final adjudication of variant and generation of .vcf file

\*consider Freebayes as it has options that enable all NRAs to be identified

\*\* consider VEP if available as a stand-alone module