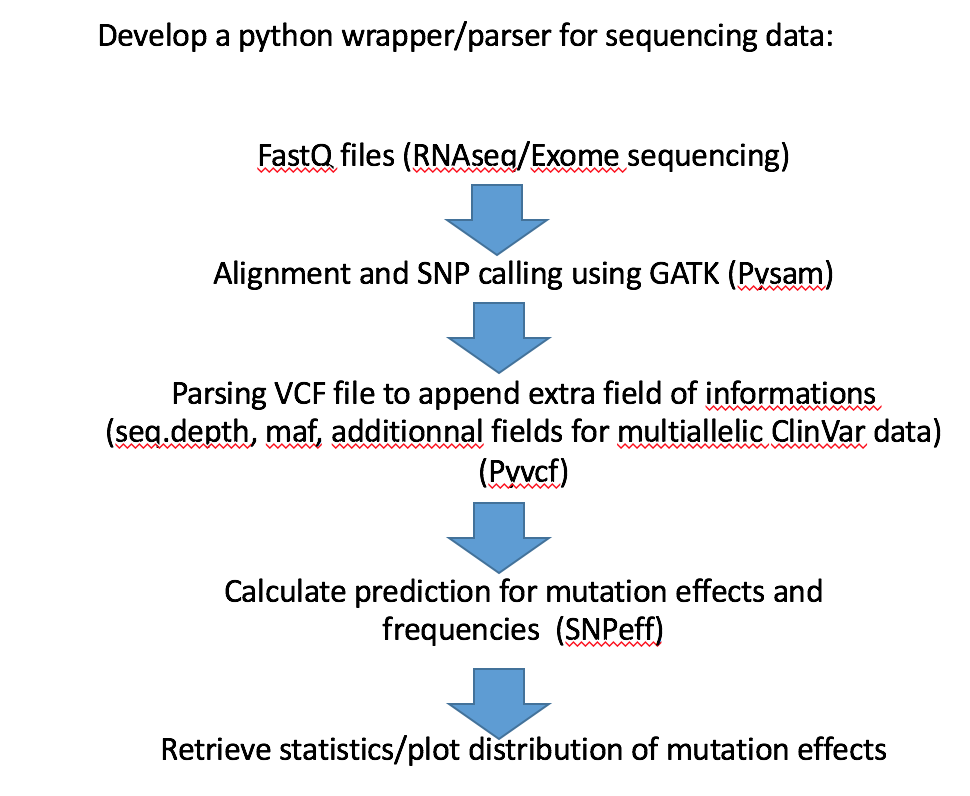
MAPPING PIPELINE FOR DRUG RESISTANCE PROJECT PROPOSAL

Jason Miller, Chingiz Underbayev, Guillaume Bauchet

Cancer is a disease driven by evolutionary selection on somatic genetic and epigenetic alterations. Despite treatment with effective drug part of the patient’s cohort resist the therapy and progress. Using whole exome sequencing data define the variants responsible for drug resistance.



INPUT

Fastq files of 13 patients with resistance to drug + 13 control patients (26 fastq files, could be less if too much). Or is there an alternative –smaller- set we could use to perform the testing? (like Ecoli or other small org., to be defined)

<https://www.ncbi.nlm.nih.gov/sra?linkname=bioproject_sra_all&from_uid=394578>

PMID:28667884

OUTPUT

1. Vcf files including variant + mutation predictions
2. Differential analysis type for variant frequency between case/control (similar to differential expression for RNAseq data) and related gene annotation
3. Graphical distribution plot of mutations along chromosomes

LEARNING GOALS

Wrap different tools and outputs types (graph, file) and do vcf file adhoc formatting using python.