**Dataset**

I worked with three sets of targeted re-sequencing data. Each comprises two FASTQ files containing paired-end raw reads from an Illumina MiSeq sequencer.

**Alignment**

After checking the quality of the datasets (FastQC), I mapped them to a reference genome called hg19 (Map with BWA-MEM), turning the FASTQ files into BAM files. After labelling each set of aligned sequences by the source (AddOrReplaceReadGroups), I merged the three BAM files into one (MergeSamFiles). Then, I removed the low-quality (Filter, mapQuality >= 20) and duplicated maps (MarkDuplicates), and soft-clipped the maps that go beyond their reference sequences, setting MAPQ to 0 for every unmapped read (CleanSam).

**Variant calling**

I identified every polymorphic site, generating a VCF file (FreeBayes). To select the polymorphic sites that appear to have strong support for polymorphism, I only retained the sites whose chances of being a false positive are below 1 in 10,000 (VCFfilter, QUAL > 40), leading to a smaller VCF file: the default henceforth.

Using VCFfilter again, I found out the genotypes of the retained sites: single nucleotide variants (TYPE = snp), deletion variants (TYPE = del), insertion variants (TYPE = ins), multi-nucleotide variants (TYPE = mnp), and complex variants like composite insertion and substitution events (TYPE = complex). I also identified the sites with multiple alternative alleles in the default VCF file (Filter, len(c5.split(','))>1).

**Annotation**

I annotated the default VCF file (ANNOVAR Annotate VCF), counted the number of polymorphic sites in each gene (Group, c7: count[c7]), and sorted the genes in a descending order of counts (Sort).

**Results**

I found 2,295 single nucleotide variants, 131 deletions, 125 insertions, and 156 multi-nucleotide variants. I found 23 variants with multiple alternative alleles. The five genes with the most polymorphic sites are RBFOX1 (167), CACNA1H (67), ABAT (47), PKD1 (41), and LMF1 (36).

Galaxy workflow: https://usegalaxy.org/u/kywertheim/w/genomic-data-science-with-galaxy