**Polymorphic Site Detection**

Step 1: Load files containing the paired end reads for father, mother, daughter trio (NA12877, NA12878, and NA12880 respectively) into a new galaxy history.

Step 2: Run quality control checks on raw data files using FastQC tool.

Step 3: Map sets of reads to the reference genome “hg19” using Map with BWA-MEM tool.

Step 4: Add labels using AddorReplaceReadGroups tool.

Step 5: Merge files using MergeSamFiles tool.

Step 6: Filter for high quality reads, MarkDuplicates and CleanSam.

Step 7: Detect genetic variants with FreeBayes tool.

Step 10: Filter to only retain sites where the chance of false positive call is 1 in 10,000 or better using VCFfilter. The resulting VCF file contains all the filtered variants.

Step 11: Use VCF filter to determine specific types of polymorphic sites with QUAL > 40.

1. SNP’s = 2,250
2. Insertions = 123
3. Deletions = 128
4. Multi-nucleotide polymorphisms = 5
5. Multiple alternate alleles = 16

Step 12: Select individual samples using VCFselectSamples tool.

Step 13: Annotate variants in each sample using ANNOVARAnotateVCF tool.

Step 14: Concatenate the 3 annotated variants datasets into a single dataset using Concatenate datasets (cat) tool.

Step 15: Group and count on col 7 (Gene.refGene).

Step 16: Sort (descending order) to determine the 5 genes with the largest number of polymorphic sites.

1. RBFOX1 - 510
2. CACNA1H - 165
3. ABAT - 151
4. USP7 - 124
5. ADCY9 - 108

**Conclusion**: We mapped the paired reads for the father, mother and child trio against the reference genome (hg19) and used the galaxy tools listed above to annotate, preprocess our data, and filter out polymorphic sites with a chance of false positive call less than 1 in 10,000. We discovered 2,250 SNP variants, 123 insertion variants, 128 deletion variants, 5 multi-nucleotide variants and 16 multiple alternate alleles variants. We determined the 5 genes that have the largest number of polymorphic sites.