

Lab Exercise 12

1)

The file **brca1_variants.csv** contains known variants in the BRCA1 gene and their pathogenicity evaluation. The file **ALL.chr17.integrated_phase1_v3.20101123.snps_indels_svsvs.genotypes.41100000-41300000.vcf.gz** is part of phase-1 of the 1000 Genomes Project that contains a 200,000 bp region of chromosome 17 surrounding the BRCA1 gene.

Find all the individuals in phase-1 of the 1000 Genomes Project that have a pathogenic BRCA1 variant.

Note that the genomic locations in the provided CSV are 0-based, but the VCF is 1-based.