



Peer-graded Assignment: Course Project

You passed!

Congratulations. You earned 15.5 / 16 points. Review the feedback below and continue the course when you are ready. You can also help more peers by reviewing their submissions.

[Review assignments](#)[Instructions](#)[My submission](#)

Identifying polymorphic sites

[Discussions](#)

Submitted on August 24, 2020

[Shareable Link](#)

PROMPT

Use the text box below to submit a short write-up describing your results including the information requested above. (Maximum 300 words)

Sequences are mapped to hg19 uses BWAM-MEM tool. Merged with MergeSamFiles and Filtered, MarkDuplicated, Clean Sam.

There's 2666 polymorphic sites with 1 in 10000 false positive in vcf qual field.

1. Single nucleotide variants: 2328
2. insertion/deletion variants: 277
3. multi-nucleotide variants: 23
4. variants with alternate alleles: 62

AnnotateVCF tool and Group then Sort identified top 5 genes with largest sites. Listed below:

1. RBFOX1: 161 Sites
2. CACNA1H: 55 Sites
3. USP7: 48 Sites
4. ABAT 44 Sites
5. CLCN7: 39 Sites