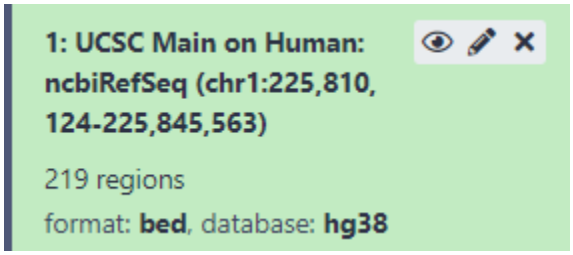
Obtain the number of common SNPs in the exonic regions of EPHX1 gene (NM_001291163).

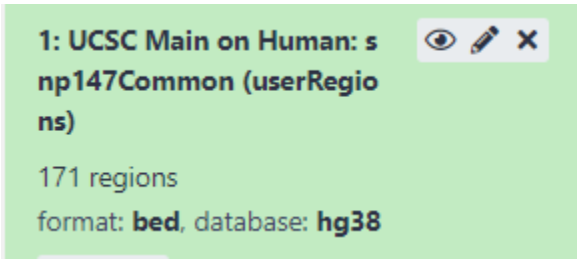
a. Used Galaxy to retrieve all RefSeq Genes (hg38) from the UCSC Main table browser in the region for the EPHX1 (NM_001291163) gene with results in BED format. Identified coding exons.

219 coding exons were identified.



b. Retrieved Common SNPs (147) from the same region and identified regions.

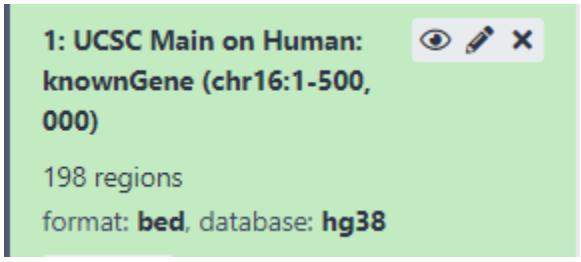
171 regions identified.



Part 2

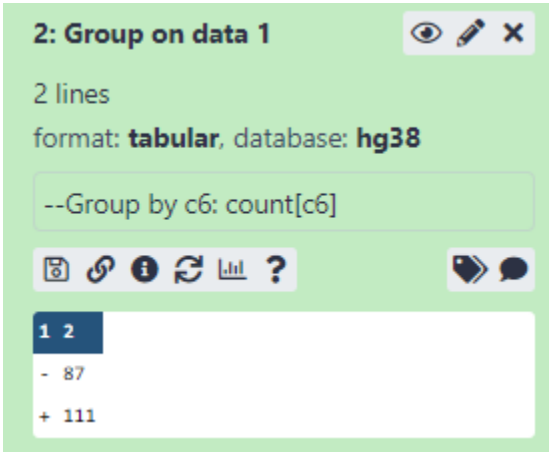
a. Used Galaxy to query for hg38 genes in ENCODE region ENm008 (chr16:1-500000) from the UCSC Main table browser at UCSC (group: Genes and Gene Predictions, track: GENCODE v32 or newer, output format: BED). Identified genes.

198 genes identified.

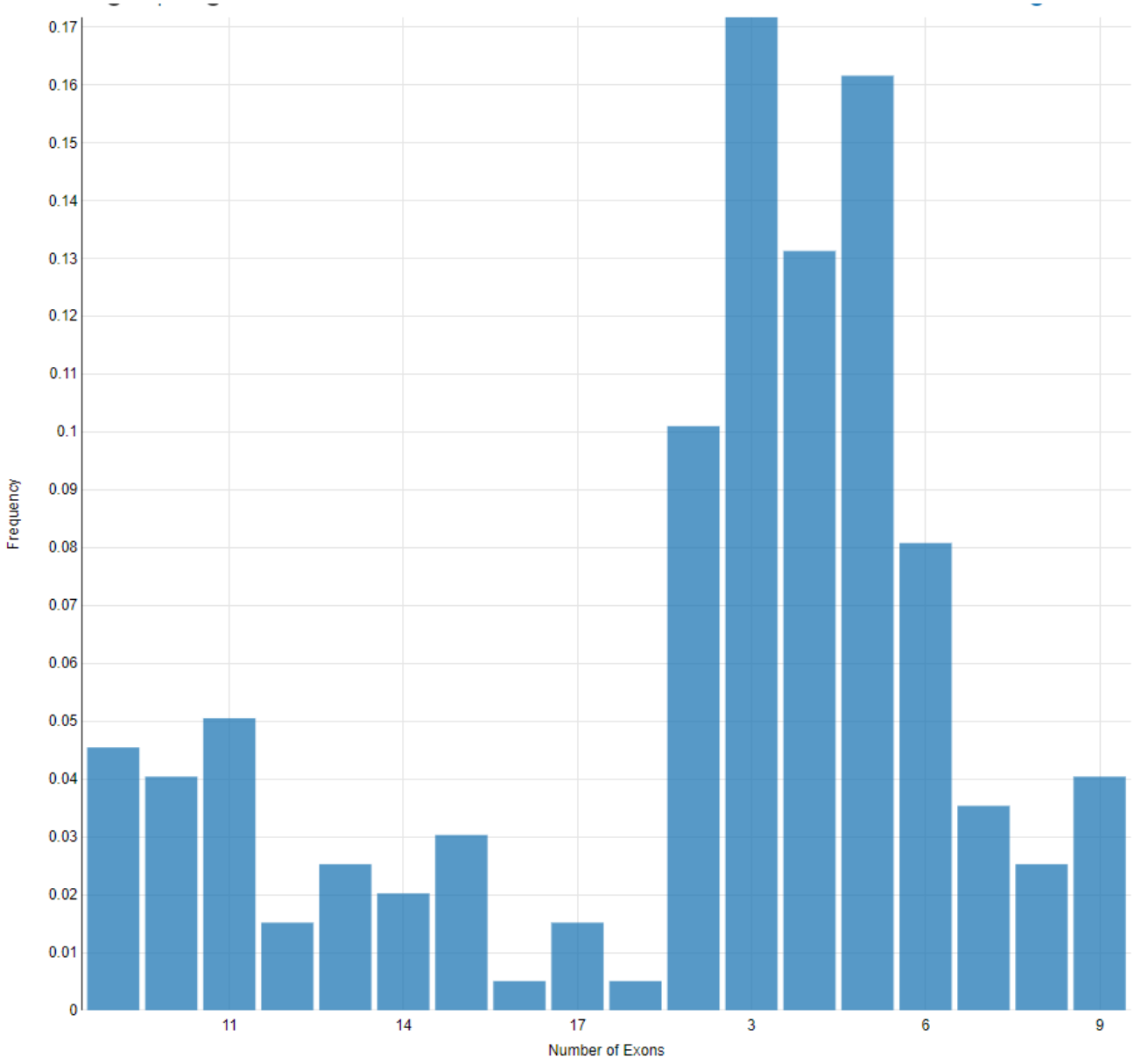


b. Determined how many genes in this output are on the plus strand and the minus strand. Using the group tool making sure to group on the correct column of the BED file and using the "Count" operation. Identified how many genes are on the plus strand and how many on the minus strand.

87 on - strand
111 on + strand



c. Generated a histogram of the number of exons in each gene.



Part 3

Used IGV for hg19, to load dbSNP 1.4.7 or newer (i.e. Available Datasets > Annotations > Variation and Repeats > dbSNP 1.4.7) and an exome sequencing track from the 1000 Genomes project (1000 Genomes > Alignments > GBR > exome > HG00096 exome). Went to the EPHX1 gene and zoomed in on the exon #4.

a. Identified how many SNPs overlap this exon and what the SNP IDs are.

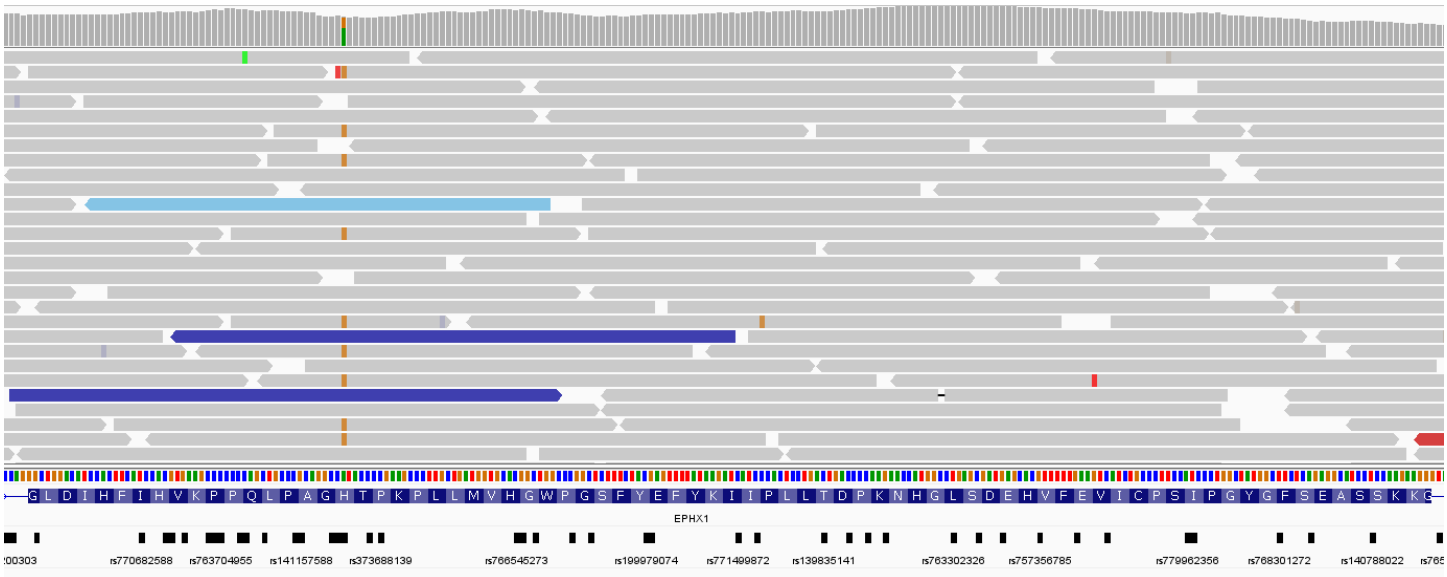
42 SNPs overlap with exon #4. Their SNP IDs are:

rs746998584, rs770682588, rs112721617....

b. Identified at which SNP this individual appear to be heterozygous. Identified the sequence count for each nucleotide at this position

At SNP rs2234922 it appears heterozygous. The sequence count is 57 A's, and 34 G's. The count is zero for C, T, and N.

c. Image from IGV, zoomed in on but showing all of the exon #4 including the SNP and exome tracks.



Part 5

Found the human CACNA1A gene in NCBI ClinVar. Filter to limit results to Variation type: Deletion.

a. Found how many total variations are listed after filtering:

103 variations are of deletion.

b. Identified how many variants from part a are pathogenic?

56 variants from the previous answer are also pathogenic.

c. Found the CACNA1A gene in Ensembl and looked for structural variants.

There are 1150 structural variants.

d. Found the CACNA1A gene in the NCBI Variation Viewer. Filtered for dbVar and CNV.

455 total CNVs.

e. Identified what is the most typical pathogenic condition for the CACNA1A gene variation from ClinVar.

Episodic ataxia type 2 is the most typical, with the condition occurring 33 times.

f. (0.25 pts) For the previous part5e, below is a clinical treatment options available.

The most common and successful treatment option appears to be acetazolamide.

Source: Strupp, M., Kalla, R., Dichgans, M., Freilinger, T., Glasauer, S., & Brandt, T. (2004). Treatment of episodic ataxia type 2 with the potassium channel blocker 4-aminopyridine. Neurology, 62(9), 1623-1625.