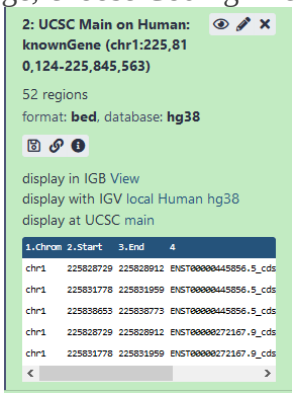
**Alex Gilson**
**AS.410.635.81**
**Genomics Spring 2021**
**Units 5-6 Graded Homework**

## Part 1

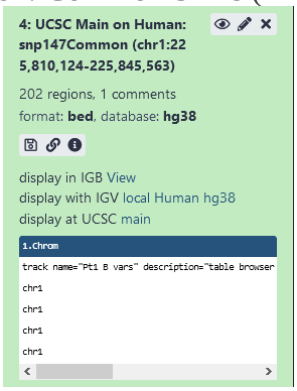
We aim to get the number of common SNPs in the exonic regions of EPHX1 gene (NM_001291163).

a. (0.5 pts) Use Galaxy to retrieve all RefSeq Genes (hg38) from the UCSC Main table browser in the region for the EPHX1 (NM_001291163) gene (type in the RefSeq ID, click lookup, then click on the RefSeq ID) and output the results in BED format. On the next page, choose Coding Exons. How many coding exons were identified?

a. 

52 coding exons were ID'd by USCS table browser and shown on Galaxy.

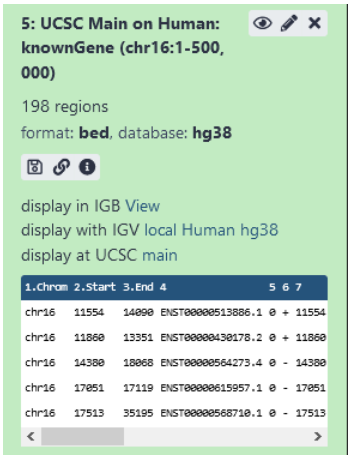
b. (0.75 pts) Retrieve Common SNPs (147) from the same region. You may use group: Variation: Common SNPs (147). How many regions were identified?

a. 

202 Common SNPs were ID'd.

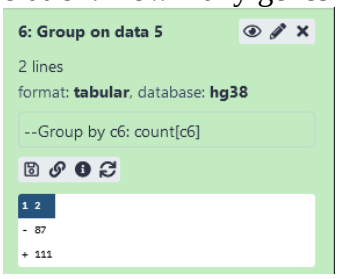
## Part 2

a. (0.25 pts) Use 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). On the next page, choose Whole Gene. How many genes were identified?

- 

198 total genes were identified.

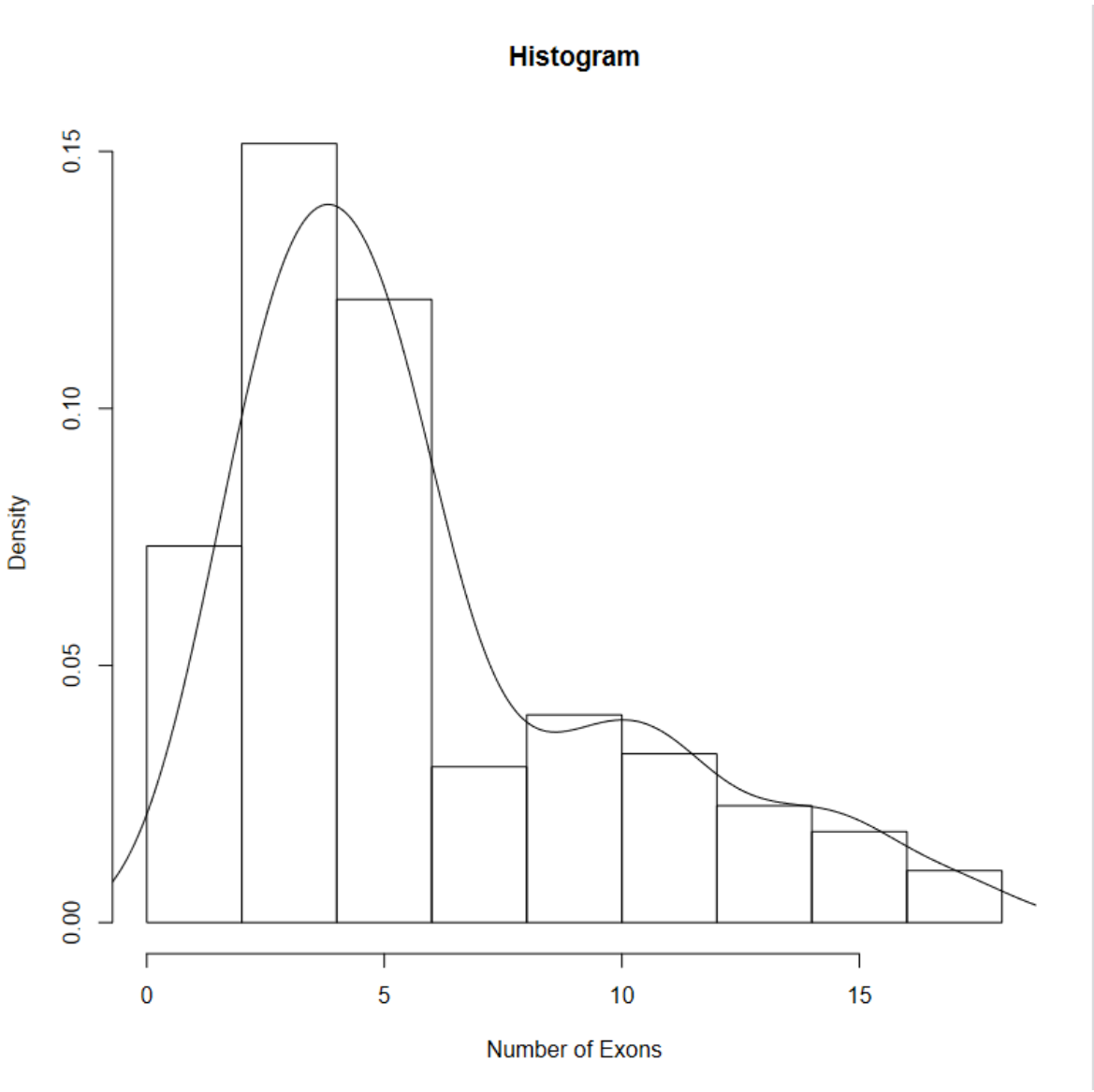
b. (0.25 pts) Determine how many genes in this output are on the plus strand and the minus strand. Use the *Group* tool making sure to group on the correct column of the BED file and using the "Count" operation. How many genes are on the plus strand? the minus strand?

- 

From the Group tool, it was found that 87 genes from the search are on the minus strand and 111 are on the plus strand.

c. (0.25 pts) Draw a histogram of the number of exons in each gene. You may use any tools. Check out details on BED12 format here if you're not sure which column to use. Play around with the parameters to produce a graph that looks good. Don't forget to label the x axis! Submit the histogram.

**The "Histogram of a numeric column" function on Galaxy was used show the frequency at which genes for specific number of exons was seen in the dataset.**




**Part 3**

Using IGV for hg19, 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). Go to the EPHX1 gene and zoom in on the exon #4.

a. (0.25 pts) How many SNPs overlap this exon and what are the SNP IDs?

- Found on IGV, and retrieved the list via UCSC

| | | | | | |
|---|---|---|---|---|---|
| chr1 | 226026355 | 226026356 | rs746998584 | 0 | + |
| chr1 | 226026372 | 226026373 | rs770682588 | 0 | + |
| chr1 | 226026376 | 226026377 | rs112721617 | 0 | + |
| chr1 | 226026377 | 226026378 | rs745675416 | 0 | + |
| chr1 | 226026379 | 226026380 | rs769262345 | 0 | + |
| chr1 | 226026383 | 226026383 | rs756407271 | 0 | + |
| chr1 | 226026383 | 226026384 | rs775238551 | 0 | + |
| chr1 | 226026383 | 226026384 | rs764460024 | 0 | + |
| chr1 | 226026384 | 226026385 | rs762318357 | 0 | + |
| chr1 | 226026436 | 226026437 | rs779549215 | 0 | + |
| chr1 | 226026442 | 226026443 | rs147296174 | 0 | + |
| chr1 | 226026445 | 226026446 | rs144904318 | 0 | + |
| chr1 | 226026454 | 226026455 | rs199979074 | 0 | + |
| chr1 | 226026455 | 226026456 | rs761426131 | 0 | + |
| chr1 | 226026469 | 226026470 | rs771499872 | 0 | + |
| chr1 | 226026472 | 226026473 | rs777312225 | 0 | + |
| chr1 | 226026483 | 226026484 | rs139835141 | 0 | + |
| chr1 | 226026487 | 226026488 | rs754290186 | 0 | + |
| chr1 | 226026490 | 226026491 | rs765684911 | 0 | + |
| chr1 | 226026493 | 226026494 | rs376877493 | 0 | + |
| chr1 | 226026504 | 226026505 | rs763302326 | 0 | + |
| chr1 | 226026385 | 226026386 | rs763704955 | 0 | + |
| chr1 | 226026388 | 226026389 | rs773821078 | 0 | + |
| chr1 | 226026389 | 226026390 | rs761335756 | 0 | + |
| chr1 | 226026392 | 226026393 | rs754199890 | 0 | + |
| chr1 | 226026397 | 226026398 | rs755378466 | 0 | + |
| chr1 | 226026398 | 226026399 | rs141157588 | 0 | + |
| chr1 | 226026403 | 226026404 | rs752988508 | 0 | + |
| chr1 | 226026404 | 226026405 | rs55784606 | 0 | + |
| chr1 | 226026405 | 226026406 | rs2234922 | 0 | + |
| chr1 | 226026409 | 226026410 | rs761017131 | 0 | + |
| chr1 | 226026411 | 226026412 | rs373688139 | 0 | + |
| chr1 | 226026433 | 226026434 | rs766545273 | 0 | + |
| chr1 | 226026434 | 226026435 | rs769650380 | 0 | + |
| chr1 | 226026508 | 226026509 | rs764312984 | 0 | + |
| chr1 | 226026512 | 226026513 | rs151050888 | 0 | + |
| chr1 | 226026518 | 226026519 | rs757356785 | 0 | + |
| chr1 | 226026524 | 226026525 | rs371520880 | 0 | + |
| chr1 | 226026529 | 226026530 | rs755831510 | 0 | + |
| chr1 | 226026542 | 226026543 | rs779962356 | 0 | + |
| chr1 | 226026543 | 226026544 | rs748882501 | 0 | + |
| chr1 | 226026557 | 226026558 | rs768301272 | 0 | + |
| chr1 | 226026562 | 226026563 | rs778715423 | 0 | + |
| chr1 | 226026572 | 226026573 | rs140788022 | 0 | + |

b. (0.25 pts) At which SNP(s) in part a does this individual appear to be heterozygous? What is the sequence count for each nucleotide at this(these) position(s) (**Hint**: look at the HG00096 exome Coverage track)

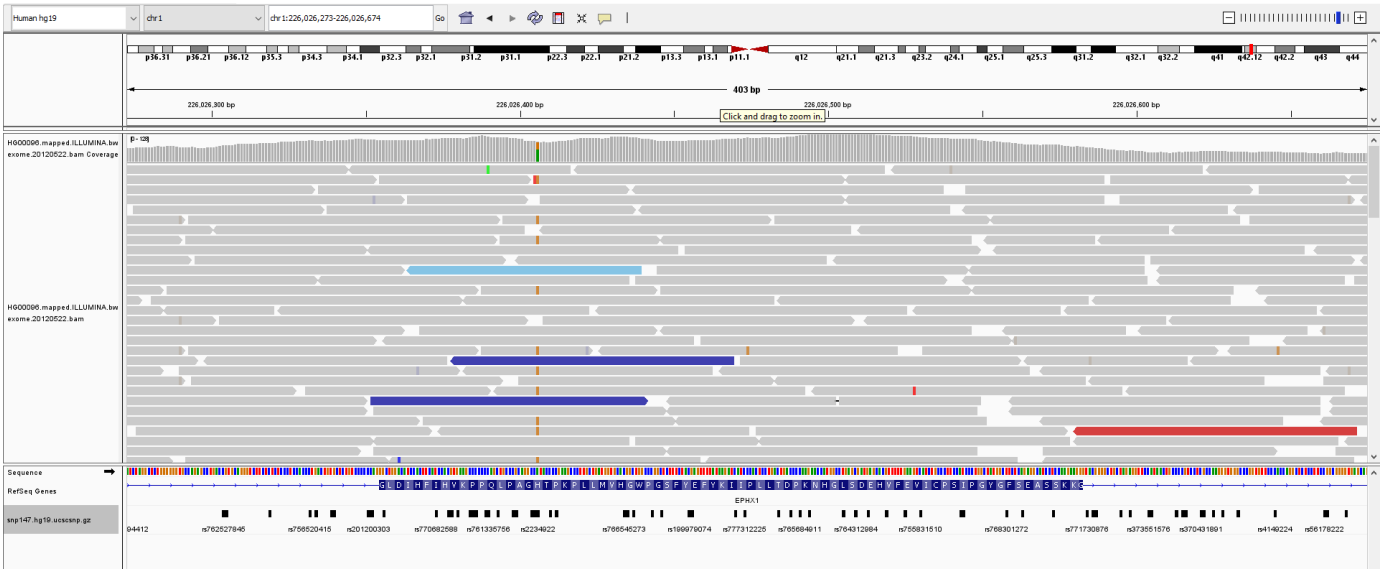
- SNP rs2234922 appears to be heterozygous at position chr1:226026406.

The sequence count is as follows:

chr1:226,026,406

Total count: 91
A : 57 (63%, 27+, 30- )
C : 0
G : 34 (37%, 20+, 14- )
T : 0
N : 0
---

c. (0.25 pts) Submit the image from IGV, zoomed in on but showing all of the exon #4 including the SNP and exome tracks.




## Part 4

a. (0.25 pts) Using IGV, load the Firehose (TCGA) data from January 28, 2016 for OV-TP (ovarian cancer) using hg19. The following four genes have been shown to be associated with CNV in some forms of ovarian cancer: PKN2, GRXCR1, PRKN (or PARK2), and PPIAL4A. In a table, qualitatively evaluate the CNV summary (minus germline) in each gene's region (*e.g.* "even blue and red", "twice as much blue as red", "overwhelmingly red").

- PKN2: Majority Red
- GRXCR1: Slightly more blue than red
- PRKN/PARK2: Majority Blue
- PPIAL4A: Completely Red(chr1:147,954,635-147,955,419)

b.(0.25 pts) Repeat part a but for Firehose (TCGA) data for BRCA-TP (breast cancer). Add CNV summaries for part b as another column in the same table as part a.

- PKN2: Slightly more blue
- GRXCR1: Majority Blue
- PRKN/PARK2: Majority Blue
- PPIAL4A: Completely Red

c. (0.25 pts) Submit the table and briefly summarize the similarities or differences between ovarian cancer and breast cancer in each of these genes in terms of CNVs. Make a third column in the table for these summaries.

| Gene | Ovarian Cancer | Breast Cancer | Comparison of Results |
|---|---|---|---|
| PKN2 | Majority Red | Slightly more Blue | Ovarian cancer shows a loss while breast cancer shows a gain, which could indicate that this gene is involved in the pathway for one of these diseases, but not the other. |
| GRXCR1 | Slightly more Blue | Majority Blue | Since both show more blue here, this gene could be implicated in both cancers, with BC showing more copy number gain and therefore more effect from the gene in the development of that cancer type, BUT could also mean that this gene plays little role in the development of these diseases. |
| PRKN | Majority Blue | Majority Blue | As both are majority blue, it is similar to GRXCR1 where the gain of CN in PRKN plays a role in the development of both OC and BC, or since it shows similarly in both, doesn't play much of a role in the difference between the two diseases. |
| PPIAL4A | Completely Red | Completely Red | Both cancers show a loss of CNV in this region of the genome, indicating that PPIAL4A is implicated in both diseases, where the loss of this gene's expression could lead to cancerous conditions. |

## Part 5

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

a. (0.25 pts) How many total variations are listed after filtering?

**103 results are shown after filtering.**

b. (0.25 pts) How many variants from part a are pathogenic?

**56 results are listed as "pathogenic".**

c. (0.25 pts) Find the CACNA1A gene in Ensembl and look for structural variants. How many structural variant entries can you find from the excel table?

**59 entries are found in the excel file.**

d. (0.25 pts) Find the CACNA1A gene in the NCBI Variation Viewer. Filter for dbVar and CNV. How many total CNVs?

**455 total CNVs are listed after filtering for those conditions on Variation Viewer.**

e. (0.25 pts) What is the most typical pathogenic condition for the CACNA1A gene variation from ClinVar?

**Episodic Ataxia type 2**

f. (0.25 pts) For the previous part5e, what clinical treatment options available? (you can use google or uptodate or epocrates or pubmed).

**According to this healthline.com article regarding types of episodic ataxia, treatment includes: anticonvulsant/antiseizure medications(acetazolamide being the most common); flunarizine and dalfampridine; physical therapy to improve strength and mobility in patients or lifestyle/diet changes.**

**https://www.healthline.com/health/episodic-ataxia#treatment**