Jaimee Beckett
Units 5-6 Graded Homework

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

      There were 219 exons identified

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

      There are 202 regions identified

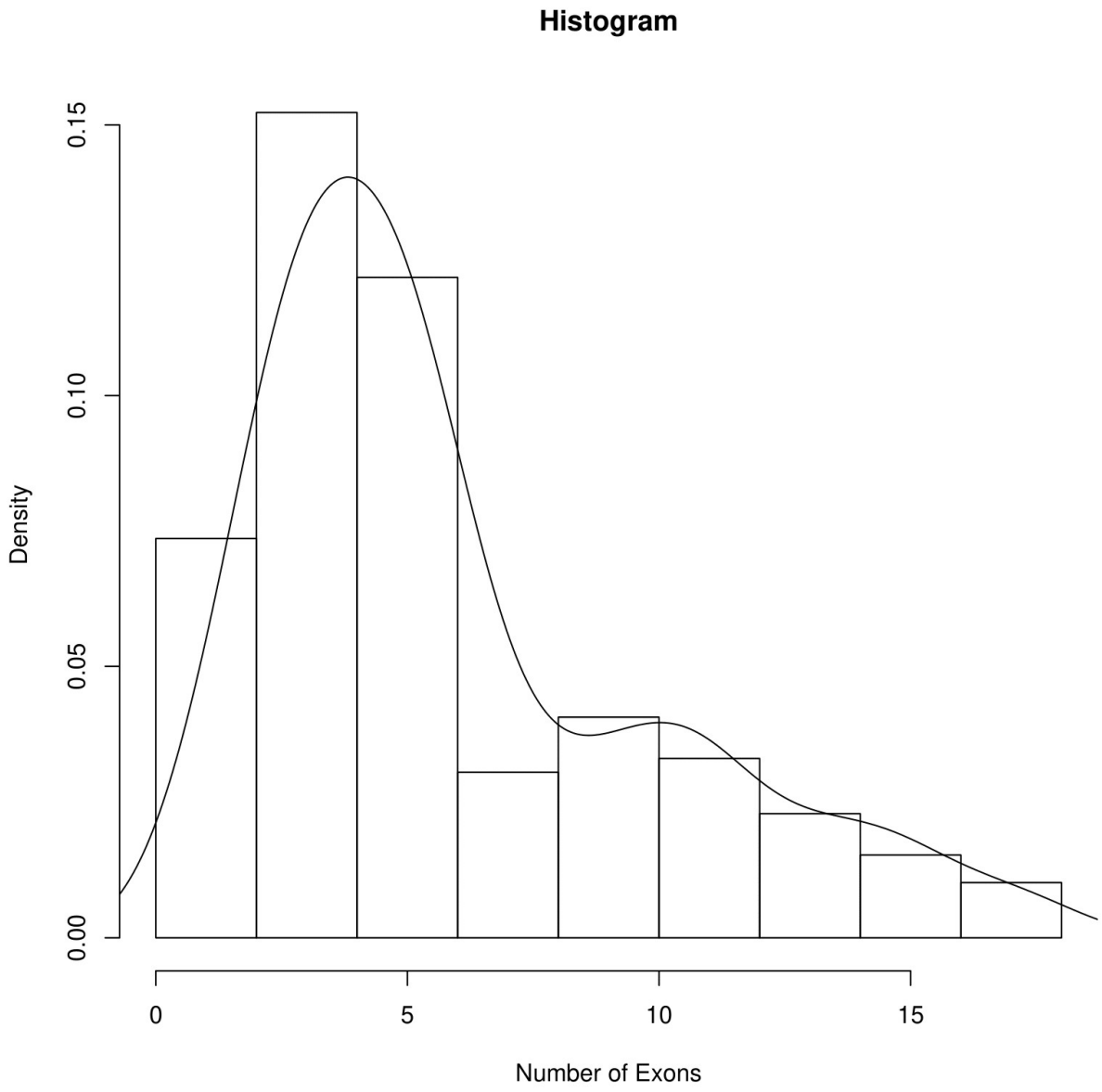
2.
   a. Use Galaxy to query for hg38 genes in ENCODE region ENm008 (chr16:1-500000) from the UCSC Main table browser at UCSC (group: Genes and Gene Predictions, track: GENCODE v32 or newer, output format: BED). On the next page, choose Whole Gene. How many genes were identified?

      There were 197 genes identified

   b. 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?

      There are 111 genes on the plus strand and 86 genes on the minus strand

   c. 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.



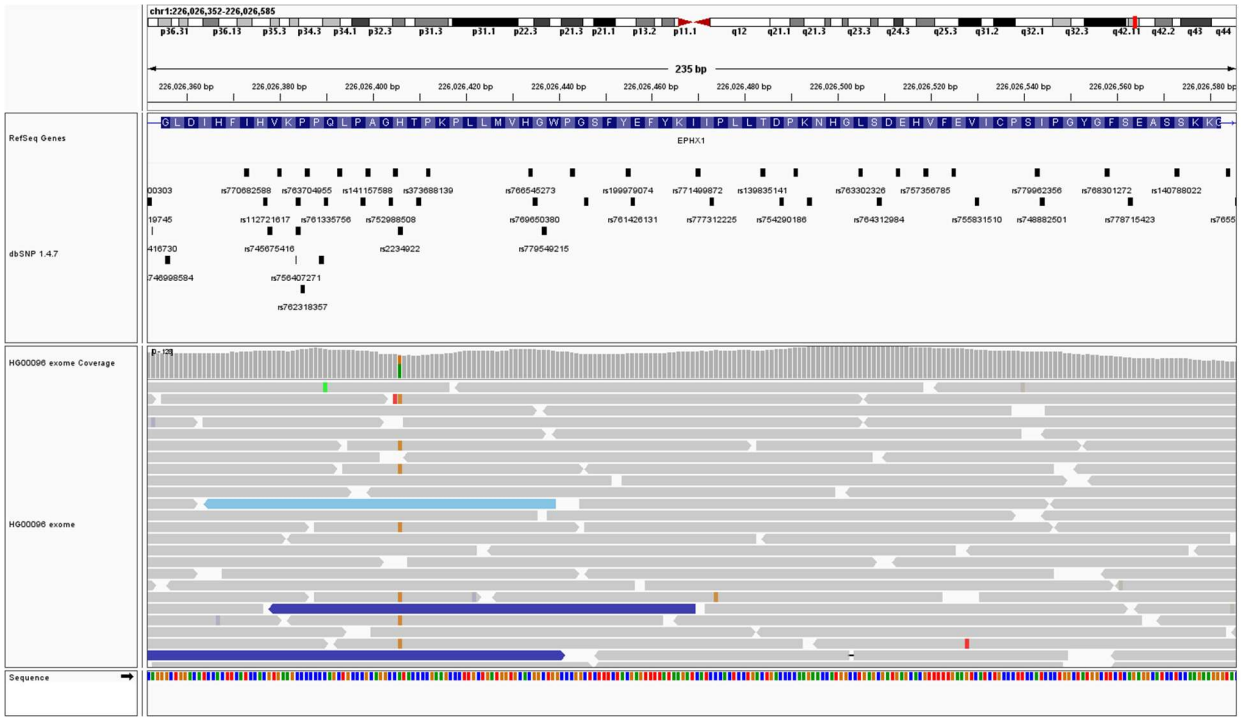
3. Using IGV for hg19, load dbSNP 1.4.7 or newer (i.e. Available Datasets > Annotations > Variation and Repeats > dbSNP 1.4.7) and an exome sequencing track from the 1000 Genomes project (1000 Genomes > Alignments > GBR > exome > HG00096 exome). Go to the EPHX1 gene and zoom in on the exon #4.
   a. How many SNPs overlap this exon and what are the SNP IDs?
      There are 44 SNPs (rs746998584, rs770682588, rs112721617, rs745675416, rs769262345, , rs764460024, rs756407271, rs762318357, rs775238551, rs773821078, rs761335756, rs755378466, rs141157588, rs763704955, rs754199890, rs55784606, rs141157588, rs752988508, rs22234922, rs761017131, rs373688139, rs766545273, rs769650380, rs779549215, rs147296174, rs144904318, rs140788022, rs778715423, rs768301272, rs748882501, rs779962356, rs755831510, rs371520880, rs757356785, rs151050888,

rs764312984, rs763302326, rs376877493, rs765684911, rs754290186, rs139835141, rs777312225, rs771499872, rs761426131)

b. 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)

rs2234922 A: 57, G: 34

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



4.

a. 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").

b. 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.

c. 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.

Most BRCA-TP patients did not have an increase or decrease in copy number for each gene whereas the opposite was true for OV-TP patients.

Gene	OV-TP	BRCA-TP	Summary
PKN2	Red: 50%, Blue: 23%	Red: 14%, Blue: 32%	OV-TP: increase in copy number for half the patients CN for BRCA-TP: 54% no change, but 32% had a decrease
GRXCR1	Red: 30%, Blue: 45%	Red: 12%, Blue: 30%	OV-TP: decrease in copy number for 45% but increase in 30% BRCA-TP: 58% no change, but 30% had a decrease
PRKN	Red: 17%, Blue: 64%	Red: 13%, Blue: 36%	OV-TP: decrease in copy number for 64% BRCA-TP: 51% no change, but 36% had a decrease
PPIAL4A	Red: 60%, Blue: 2%	Red: 28%, Blue: 3%	OV-TP: increase in copy number for 60% BRCA-TP: 69% no change, but 28% had an increase

5. Find the human CACNA1A gene in NCBI ClinVar. Filter to limit results to Variation type: Deletion.
   a. How many total variations are listed after filtering?

      118

   b. How many variants from part a are pathogenic?

      64

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

      1154

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

0

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

Epileptic encephalopathy, early infantile, 42, Episodic ataxia type 2

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

Anti-epileptic medications such as Sulthiame