Julie Garcia
Unit 5-6 Graded Homework
February 16, 2018

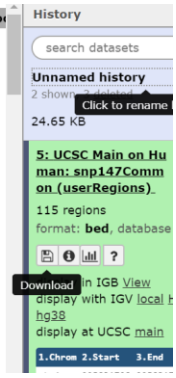
Part 1

a. I used the Main Table Browser to get the following data. Here is a screenshot of the results of the query in BED format in Galaxy. The results show 183 coding exons.

Chrom	Start	End	Name	Score	Strand	ThickSta
chr1	225828729	225828912	NM_001136018.3_cds_1_0_chr1_225828730_f	0	+	
chr1	225831778	225831959	NM_001136018.3_cds_2_0_chr1_225831779_f	0	+	
chr1	225838653	225838881	NM_001136018.3_cds_3_0_chr1_225838654_f	0	+	
chr1	225839216	225839346	NM_001136018.3_cds_4_0_chr1_225839217_f	0	+	
chr1	225839828	225840037	NM_001136018.3_cds_5_0_chr1_225839829_f	0	+	
chr1	225842365	225842474	NM_001136018.3_cds_6_0_chr1_225842366_f	0	+	
chr1	225844497	225844623	NM_001136018.3_cds_7_0_chr1_225844498_f	0	+	
chr1	225845145	225845347	NM_001136018.3_cds_8_0_chr1_225845146_f	0	+	
chr1	225828729	225828912	NM_001291163.1_cds_1_0_chr1_225828730_f	0	+	
chr1	225831778	225831959	NM_001291163.1_cds_2_0_chr1_225831779_f	0	+	
chr1	225838653	225838881	NM_001291163.1_cds_3_0_chr1_225838654_f	0	+	
chr1	225839216	225839346	NM_001291163.1_cds_4_0_chr1_225839217_f	0	+	
chr1	225839828	225840037	NM_001291163.1_cds_5_0_chr1_225839829_f	0	+	
chr1	225842365	225842474	NM_001291163.1_cds_6_0_chr1_225842366_f	0	+	
chr1	225844497	225844623	NM_001291163.1_cds_7_0_chr1_225844498_f	0	+	
chr1	225845145	225845347	NM_001291163.1_cds_8_0_chr1_225845146_f	0	+	
chr1	225828729	225828912	NM_000120.3_cds_1_0_chr1_225828730_f	0	+	
chr1	225831778	225831959	NM_000120.3_cds_2_0_chr1_225831779_f	0	+	
chr1	225838653	225838881	NM_000120.3_cds_3_0_chr1_225838654_f	0	+	
chr1	225839216	225839346	NM_000120.3_cds_4_0_chr1_225839217_f	0	+	
chr1	225839828	225840037	NM_000120.3_cds_5_0_chr1_225839829_f	0	+	
chr1	225842365	225842474	NM_000120.3_cds_6_0_chr1_225842366_f	0	+	

b. I downloaded the above bed file (which is attached to my submission) and filtered out the last three columns. I loaded this into the Table browser as the regions to search for Common SNPs. The following is the output SNPS regions that are common to the exon areas of the EPHX1 gene. There are 115 regions. I also attached this download be file with my submission.

Chrom	Start	End	Name	Score	Strand	ThickStart	ThickEnd	ItemRGB	Blo
chr1	225831798	225831799	rs2234698	0	+				
chr1	225831884	225831885	rs4986931	0	+				
chr1	225831931	225831932	rs1051740	0	+				
chr1	225831951	225831952	rs1131873	0	+				
chr1	225838704	225838705	rs2234922	0	+				
chr1	225839846	225839847	rs34143170	0	+				
chr1	225839957	225839958	rs2292568	0	+				
chr1	225844527	225844528	rs1051741	0	+				
chr1	225845226	225845227	rs4149229	0	+				
chr1	225845328	225845329	rs4149230	0	+				
chr1	225831798	225831799	rs2234698	0	+				
chr1	225831884	225831885	rs4986931	0	+				
chr1	225831931	225831932	rs1051740	0	+				
chr1	225831951	225831952	rs1131873	0	+				
chr1	225838704	225838705	rs2234922	0	+				
chr1	225839846	225839847	rs34143170	0	+				
chr1	225839957	225839958	rs2292568	0	+				



Part 2

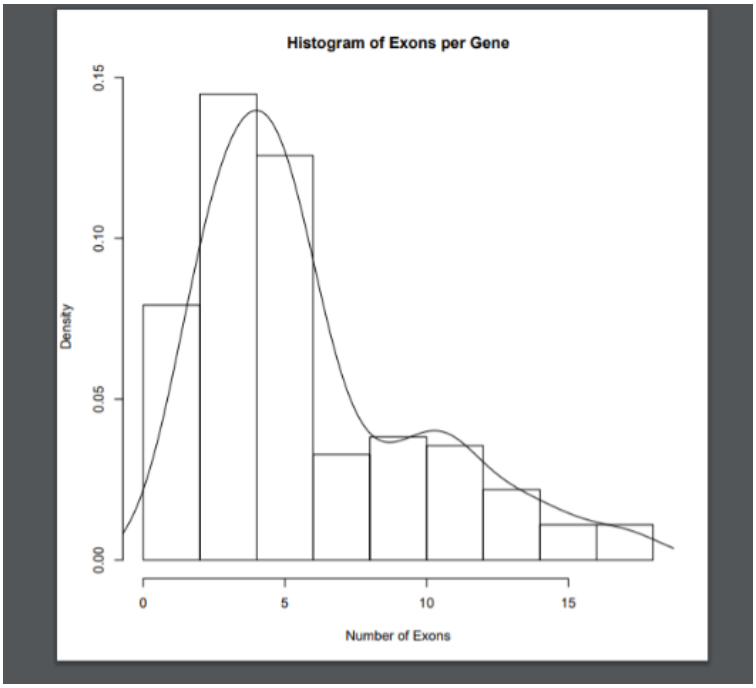
a. I used Galaxy and the Main Table Browser to query for the ENCODE region ENm008 and the results are posted below. It came up with 183 know genes.

Chrom	Start	End	Name	Score	Strand	ThickStart	ThickEnd	ItemRGB	BlockCount
chr16	17051	17119	uc032dmn.1	0	-	17051	17051	0	1
chr16	17513	35195	uc059ofm.1	0	-	17513	17513	0	4
chr16	22909	25123	uc002cfh.4	0	+	22909	22909	0	2
chr16	46406	53628	uc002cfi.3	0	-	47429	53586	0	3
chr16	47235	52142	uc059ofn.1	0	-	47235	47235	0	2
chr16	47734	48007	uc059ofo.1	0	+	47734	47734	0	1
chr16	53009	57669	uc002cfj.5	0	+	54016	57143	0	5
chr16	53828	57668	uc059ofp.1	0	+	53989	57143	0	5
chr16	53886	57372	uc059ofq.1	0	+	54016	56360	0	6
chr16	53898	55871	uc059ofr.1	0	+	53898	53898	0	3
chr16	53898	56613	uc059ofs.1	0	+	54016	56613	0	4
chr16	54959	57444	uc059oft.1	0	+	54959	54959	0	4
chr16	55494	57453	uc059ofu.1	0	+	55494	55494	0	3
chr16	58058	72631	uc002cfl.4	0	-	58339	65015	0	18
chr16	58061	65039	uc059ofv.1	0	-	62883	65015	0	17
chr16	58136	59762	uc059ofw.1	0	-	58339	59762	0	5
chr16	58973	60193	uc059ofx.1	0	-	58973	58973	0	3
chr16	59232	60680	uc059ofy.1	0	-	59232	59232	0	4
chr16	61424	62206	uc059ofz.1	0	-	61424	61424	0	3
chr16	62972	65037	uc059oga.1	0	-	63903	65015	0	4

b. I used the Group tool (under Join, Subtract and Group) to group by strand. The results are shown below. There are 99 genes on the plus strand and 84 genes on the minus strand, equaling 183 total genes as stated above. I also attached these results in my submission.

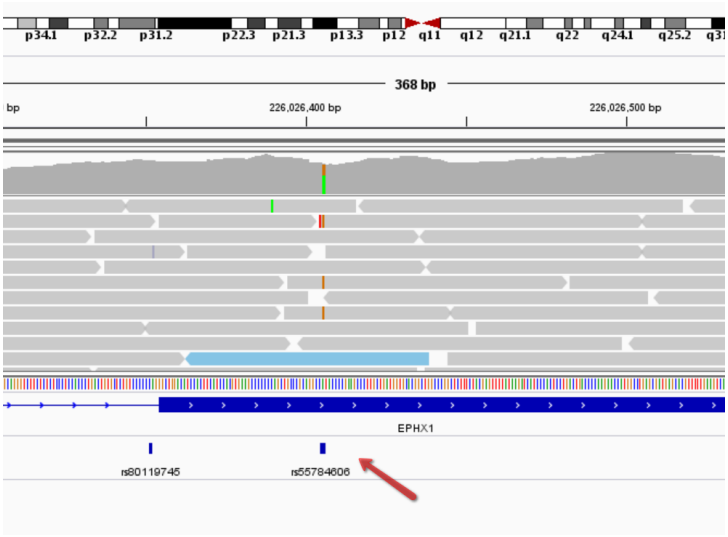
1	2
-	84
+	99

c. I used the histogram tool to create the following plot of number of exons per gene. I also attached the pdf of the histogram to my submission.

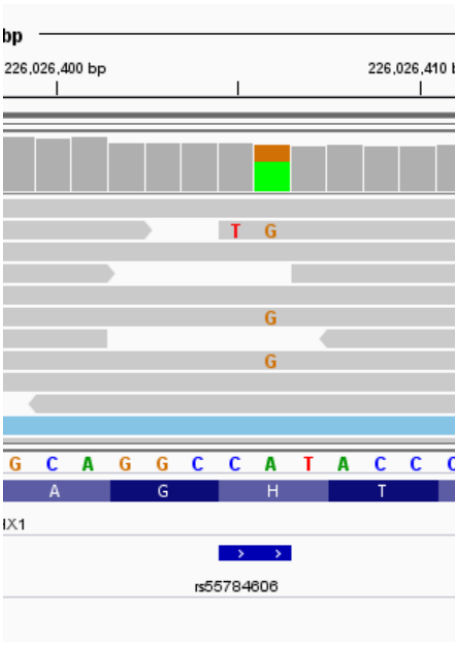


Part 3

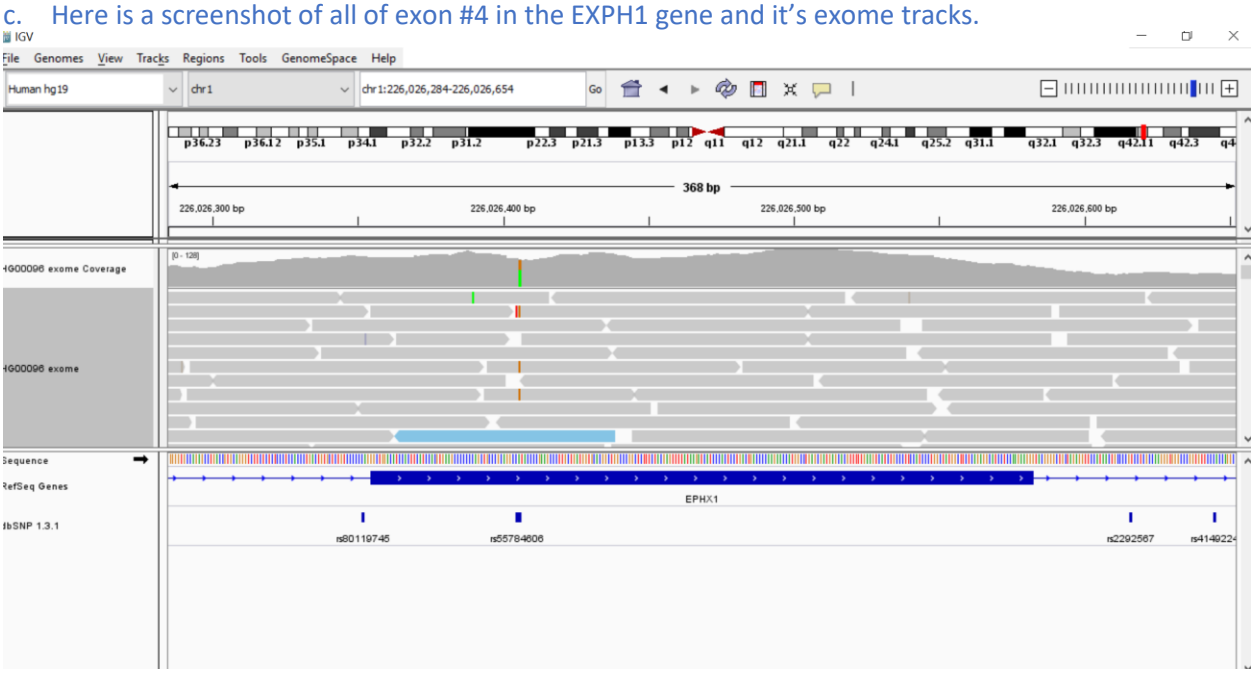
a. I loaded the EXPH1 gene in IGV and zoomed in on exon #4 and found only one SNP that is overlapping with the exon as shown below. The SNP number is rs55784606.



b. The same SNP rs55784606 appears to be heternozygous, with A:57, G:34.



c. Here is a screenshot of all of exon #4 in the EXPH1 gene and it's exome tracks.



Part 4

a. & b.

Gene	Red OV-TP	Blue OV-TP	Summary for OV-TP Dataset	Red BRCA-TP	Blue BRCA-TP	Summary for BRCA-TP	Overall Summary
PKN2	50%	23%	Twice as much red than blue, which means there are many more tumor samples with increased copy numbers in this gene.	14%	32%	Twice as many blue than red, but many white, which means decrease in copy number is more common.	In this gene, an increase in copy number seems to be much more common, while a decrease in copy number for this gene seems to be more common in breast cancer.
GRXCR1	30%	45%	Slightly less red than blue, which means there were more patient tumors with decreased copy numbers in this gene.	12%	30%	More than twice as many blue than red, but many white which means decrease in copy number is more common.	In this gene, decrease in copy number seems to be more common, more-so in breast cancer than ovarian cancer. In breast cancer there is a large amount of samples with no copy number variation.
PRKN (or PARK2)	17%	64%	Overwhelmingly blue, which means a decrease in copy number is more common in this gene for this tumor type.	13%	36%	Almost three times more blue than red, which means decrease in copy number is more common.	For both breast cancer and ovarian cancer, a decrease in copy number variation seems to be more common in this gene.
PPIAL4A	60%	2%	Overwhelmingly red, which means an increase in copy number is much more common.	28%	3%	Overwhelmingly red, which means increase in copy number is much more common in this gene.	For both breast cancer and ovarian cancer an increase in copy number variation seems to be much more common in this gene. A decrease in copy number variation is very rare.

Part 5

a. I searched for the human CACNA1A gene in NCBI ClinVar. After filtering for deletion, there are 56 variations listed.
b. 32 of these variations are pathogenic.
c. I searched for the same gene in Ensembl. I attached the downloaded csv to my submission. The table shows 991 structural variants.
d. I searched for this gene again in NCBI Variation Viewer, filtered by dbVar and CNV. There were 264 CNVs.
e. In ClinVar, the most typical pathogenic condition was Episodic ataxia type 2.
f. The most common treatment for Episodic ataxia type 2 is the drug Acetazolamide, which can be used to control the disorder, but is not a cure.

References:

1. Spacey S. Episodic Ataxia Type 2. 2003 Feb 24 [Updated 2015 Oct 15]. In: Adam MP, Ardinger HH, Pagon RA, et al., editors. GeneReviews® [Internet]. Seattle (WA): University of Washington, Seattle; 1993-2018. Available from: https://www.ncbi.nlm.nih.gov/books/NBK1501/