

Part 1

Working with two bed files: `hs_5q31_exons.bed` and `hs_5q31_cpg.bed`. The first file contains exons in chromosomal region 5q31. The second file contains CpG regions in the same chromosomal region. Used Galaxy to answer the following.

How many unique exons intersect a CpG region?

542

How many unique exons do not intersect CpG regions?

4,296

How many unique CpG regions intersect an exon?

200 regions

How many unique CpG regions do not intersect exons?

45 regions

Submit a Galaxy workflow showing all the steps used to answer parts a-d (HINT: In history panel, click on the cog > “Extract Workflow” > select only the steps that contribute to answering the HW problems > Name workflow something meaningful > “Create workflow”. Then click “Workflow” at the top of Galaxy, click on your workflow name > “Share or Download”. Submit the .ga file).

Submitted in Data folder

Part 2

Used the UCSC Table Browser to retrieve three ENCODE tracks from region `chr5:134000000-134250000` of hg38.

DNase HS (hypersensitivity), table HUVEC Pk (output a BED file)

Layered H3K4me3, HUVEC Cells (output a WIG (data points) file)

Any other ENCODE track of your choice (output appropriate file type) (**Layered H3K27Ac**)

Using IGV, loaded all three tracks to hg38.

Submitted a screenshot that includes the TCF7 gene and the 8 closest DHS regions.



For the eight DHS regions that are closest to the TCF7 gene, described what genomic features they are closest to or overlap (e.g. Located in intron. Located upstream of TSS.)

The first two DHS regions, p140458 and p140459, are located upstream of TSS. The third, p140460 starts upstream of the TSS but partially overlaps with the first exon of TCF7. The next three DHS regions (p140461, p140462, and p140463) overlap with an exon of TCF7. p140464 is located in an intron of TCF7, and lastly, p140465 is located in the last exon of TCF7.

Part 3

Go to YeastMine and click “Templates” in the top menu. Find the template for retrieving genes annotated to a specific GO (gene ontology) ID. Find all genes associated with GO:0006623 (protein targeting to vacuole). Export the results as a tab-separated file and submit the file.

File is included in data

Part 4

Using the attached Arabidopsis thaliana protein sequence (tair_plant.fasta):

Go to the TAIR website and, using the BLASTP tool, search for the closest match in the TAIR10 database. Click on the best-matching link. What is the ID and description of this protein?

The highest matching ID is AT5G13630.1, and it's for a protein that is involved in "plastid-to-nucleus signal transduction".

In what cellular compartment(s) does this protein function?

The protein functions in parts of the chloroplast inner membrane, plastid, and mitochondrion.

How many protein-coding gene models exist for this protein?

There exists 2 protein-coding gene models for this protein, with the second being AT5G13630.2.

Briefly describe how the protein-coding gene models differ.

The main difference is that AT5G13630.2's first exon is smaller, starting downstream of AT5G13630.1, and ending upstream of AT5G13630.1. The remaining exons are the same, with the exception of the last exon, which is smaller by ending more upstream.

Scroll down to BAR eFP browser and check "Developmental Map". And, describe in which of the tissue types shown in the graph is this protein expressed the lowest?

The tissues that it is the least expressed are in the root related tissue, including root without apex, root apex, and the root of a young seedling. It is also expressed very low in seeds, especially when dry, with a slight increase during germination.

Part 5

Load the attached hs_chr20* files to the BFX server. The hs_chr20_H3K4me3 BED file represents active H3K4me3 states on chr 20. The hs_chr20_refseq BED file represents coding exons on chr 20. Use bedtools intersect with the H3K4me3 file as the A file and the refseq file as the B file to answer the following:
How many unique H3K4me3 regions intersect a coding exon?

There are 10 unique regions that intersect a coding exon.

How many unique H3K4me3 regions do not intersect a coding exon?

There are 245 unique regions that do not intersect a coding exon.

Submit your command line code.

\$ bedtools intersect -u -a hs_chr20_H3K4me3.bed -b hs_chr20_refseq.bed

```
$ bedtools intersect -v -a hs_chr20_H3K4me3.bed -b hs_chr20_refseq.bed > nonoverlaps.bed  
$ wc -l nonoverlaps.bed
```

****Discussion topic: Create a BAM intersect file. Load it onto IGV (hg38). Find the exons that overlap the H3K4me3 regions.****

```
tair_plant.bed.txt  
hs_chr20_refseq.bed  
hs_chr20_H3K4me3.bed  
hs_5q31_exons.bed  
hs_5q31_cpg.bed  
hs_chr20_H3K4me3_for_linux.bed  
tair_plant.fasta
```