Hands-on: Genomic Analysis with Genome Browsers

Demo and Exercises 1:

1. Searching for your favorite gene in the genome browser.
    - Go to UCSC Genome Browser homepage at http://genome.ucsc.edu/.
    - Select "Mammal" under group, "Human" as the genome, and "Feb. 2009 (GRCh37/hg19)" as the assembly. Enter your favorite gene name in the search box under "search term", and click the "submit" button. In this example, we will use BMP4.
    - Under "RefSeq Genes" you should see two isoforms listed. Choose the first one.
2. How many exons are in the BMP4 transcript chosen? What's the strand orientation relative to the genome?
    - BMP4 has four exons in both isoforms. It's in the opposite orientation relative to the genome (ie. minus strand).
    - Note: mouse-over an exon to help you identify which exon you are viewing.
3. Does the RefSeq gene catalog contain the correct number of isoforms of your favorite human gene? How about Ensembl and GENCODE gene models? Is there supporting evidence?
    - Under the "Genes and Gene Predictions" group, change the mode of "RefSeq Genes" and "Ensembl Genes" track to "pack", and "ENCODE" track to "show".
    - Scroll down to the "mRNA and EST" group, and change "Human mRNAs" and "Human ESTs" to squish. Click on a "refresh" button.
    - Ensembl and RefSeq have similar number of transcripts, and GENCODE has fewer. Not all GENCODE models are supported by ESTs.
4. Zoom in to look for the start codon. You may need to change the track setting.
    - Zoom in to the first coding exon. For BMP4, it's the 3rd exon from RefSeq track.
    - To add the codon numbering, you can go to the track setting, and check the box next to "Show codon numbering". By default, RefSeq track doesn't display amino acids, you can enable it by selecting "genomic codon" next to "Color track by codons".

Demo and Exercises 2

1. Identify the transcription factors in your gene's promoter region.
    - Click on the "hide all" button to hide all tracks. Turn "RefSeq Genes" track to "pack". Under the Regulation group, click on the link of "ENCODE Regulation…", turn the "Txn Factor ChIP" to "pack", then click on "submit" button.

- Navigate to the promoter region of your favorite gene and look at nearby the regions bound by transcription factors. The darkness of the gray box is proportional to the maximum signal strength observed in any cell line contributing to the cluster. A green highlight indicates the highest scoring site of a canonical motif for the corresponding factor. Click on a TF (item description) to show this motif.

2. Are there any RNA-seq data expressed in your favorite cell/tissue? If so, which isoform is most likely to be expressed?

    - Click on the ENC RNA-seq track under Expression. Choose the source of the RNA-Seq (eg. ENCODE/Caltech) by clicking on the link itself, this should allow you to choose specific cell lines or tissues. You may need to configure the track to "auto-scale" to view correctly.

    - Alternatively, go to the ENCODE site: http://genome.ucsc.edu/ENCODE/, and click on the "Search" button on the left. Click on the Track Search button, Select "Advanced" search from the top. Pick "Cell, tissue or DNA sample" category and choose your favorite cell line. Next, next to "Experiment (Assay) type" category select "RNA-seq". After click on the "search" button, tracks will be displayed below the search boxes.

    - For alignment tracks, it's better to use visibility as "squish".

Demo and Exercises 3

1. View the peaks (bed format) derived from ChIP-Seq

    - To save time, we will chr22 part of the H3K36M3, http://jura.wi.mit.edu/bio/education/hot_topics/UCSCgenomeBrowser_Apr2016/H3k36me3.Peaks.chr22.bed, downloaded from GM12878 H3K36me3 Histone Mods by ChIP-seq Peaks from ENCODE/Broad. There are multiple ways to add custom tracks:

        I. Save the H3K36m3 peak file to your local computer, browse to it, and click on the Submit button.

        II. Open the H3K36m3 peak file, copy and paste its content in the text box under "Paste URLs or data" , then click on the Submit button.

        III. Put the URL of the H3K36m3 peak file under the text under "Paste URLs or data", then click on the Submit button. To find the URL, right click on the file link, and click on the "Copy Link Location".

        In the next page, click on the "go" button next to 'view in "Genome Browser"'.

2. Identify the RefSeq genes that could be regulated by H3K36me3
    - Go to the Table Browser under the "Tools" from top of the page
    - Choose group as "Genes and Gene Predictions", track as "RefSeq Genes", and table as "refGene".
    - Click on the "create" button next to intersection. Then, select your custom track in the top. Click on the "submit" button.
    - Multiple output format available. In this case, we can just output as "BED – browser extensible data". After click on the "get output" button, you can see the RefSeq ids associated with the peaks.