Exercises for the ENCODE data in the UCSC Genome Browser

1) Using RNA-seq data, examine expression of RNA in the vicinity of the TP53 gene. From the CSHL Long RNA-seq track, determine which strand is transcribed into Poly-A+ RNA and then found in the nuclear fraction of K562 and GM12878 cells.

Skills: Use RNA-seq data to evaluate RNA presence in a region; become aware available cell-fraction data.

2) In the region we are exploring, let's add transcription factor binding data and histone marks that are often found near active regulatory elements. Let's also determine if these histone marks are indicated in human embryonic stem cells.

Skills: Explore TFBS data; examine features associated with histone modifications; visualize cell type specific data.

For additional guidance and ways to interact with the ENCODE data, access this open access publication in PLoS Biology: http://bit.ly/plosENC

Citation: The ENCODE Project Consortium (2011) A User's Guide to the Encyclopedia of DNA Elements (ENCODE). PLoS Biol 9(4): e1001046. doi:10.1371/journal.pbio.1001046

UCSC ENCODE 2 Exercises, version 1 from OpenHelix, modified by UCSC.
Correspond to the data available in January 2014.

1) Using RNA-seq data, examine expression of RNA in the vicinity of the TP53 gene. From the CSHL Long RNA-seq track, determine which strand is transcribed into Poly-A+ RNA and then found in the nuclear fraction of K562 and GM12878 cells.

Step	Action	✓
1	Go to the UCSC Genome Browser homepage, **genome.ucsc.edu**	
2	From the blue navigation links on the left side of the page, **click the link for Genome Browser.**	
3	From the Gateway interface**, click the link that says "Click here to reset the browser user interface settings to their defaults."** This will ensure that any prior activity on the Browser has been cleared out and that everyone is starting with default settings.	
4	Choose the Human **February 2009 (hg19)** assembly. Enter the text **tp53** in the gene box. Choose the TP53 item in the list. Click **submit**.	
5	In the TP53 region on the browser, examine the features briefly. Then **click the "zoom out" 1.5x button** near the top. Assess the features again.	
6	**Click the "hide all" button** in the middle of the resulting Genome viewer page. *(We want to reduce what's in the display to reduce the burden on the servers, and to focus on our features of interest.)*	
7	**Add back 2 tracks** to the viewer: • **GENCODE Genes in "pack"** visibility (from the Genes and Gene Predictions group) • **ENC RNA-seq…in "show"** (from the Expression group) **Click a refresh button** to add these tracks back to the viewer. It may take a moment for this to load, as there is a lot of data here.	
8	RNA-seq data from multiple labs, cell lines, and experiment types are shown. Let's focus on the Long RNA-seq data. You can see there is signal across this region indicating RNA transcription in this region of the genome. But we'd like to distinguish which RNA-seq data corresponds to which genes in this region. **Return to the RNA-seq… hyperlink and click it to access individual tracks from this super-track.**	
9	Note all the RNA-seq… component tracks are in "dense" visibility at this time. **Turn all of them to "hide" except for the CSHL Long RNA-seq menu.**	
10	**Click the CSHL Long RNA-seq hyperlink. Examine the options you can set to explore this track's data.**	
11	**At the top, set the Maximum display mode to "full".** **Make these changes** on the Long RNA-seq settings: *** Set Contig view to "hide".** *** Unselect all localization checkboxes** *** Select the GM12878 cell line checkbox in the "Nucleus" column.** *** Select the K562 "Nucleus" checkbox also.** *** In the RNA Extract section, Poly-A+ should be checked. Uncheck "Total RNA." Leave the other settings in that area unchanged.** **Click "Submit" when these changes have been made.**	

12	**Observe the Plus- and Minus-strand RNAs in the image.** Note that each experiment has been done twice. Determine for yourself how well the replicates correlate. Note that the WRAP53 RNA derives from the plus strand, and the TP53 RNA from the minus strand. This will help you to orient when looking for transcription factor binding sites or other genomic features.	
13	**You can also access this view using the Sessions tool.** In the "My Data" pulldown, restore the session using "User: example" and "Session name: hg19_korea2014"	
14	**Remove the replicates in the graphic by using the configuration page again, setting "Replica rank" to "1st."**	
This exercise was inspired by the Figure 3 illustration in the ENCODE User Guide paper. See that figure legend and the accompanying text for more assessments of the data and the features in this region: **http://bit.ly/plosENC**		

2) In the region we are exploring, let's add transcription factor binding data and histone marks that are often found near active regulatory elements. Let's also determine if these histone marks are indicated in human embryonic stem cells.

Step	Action	✓
1	On the browser view that we established in exercise 1, **scroll down to the Regulation Group.**	
2	**Locate the ENCODE Regulation**… track. **Choose "show"** in the pulldown menu. **Click a "refresh"** button**.**	
3	**Examine the display.** New data appears in the viewer beneath the RNA-seq data. *Note that the Transcription Factor ChIP-seq from ENCODE track shows data blocks, but not individual transcription factors. *Note that the H3K27Ac histone mark track appears to have multiple data sets of various colors.	
4	**Return to the ENCODE Regulation menu** area**. Click the hyperlink** to look at the component tracks of this super-track.	
5	By default Txn Factor ChIP is visible in "dense" mode. **Set that menu to "full". Click the "Submit" button.**	
6	Examine the display again. Note that individual transcription factors can be identified by name using the labels on the left. Note that the letter codes near the blocks correspond to cell lines that have been used in experiments for this data. **Click some of the blocks** to note the cell lines and signal levels observed in them. Return to the viewer for the next steps.	
7	**Click the grey control button to the left of the Layered H3K27Ac** to go to the controls for that track**.**	
8	On this histone mark page, note that there are various cell line data sets, which have color codes. One of the lines is **H1-hESC**, which is a human embryonic stem cell line.	
9	**Uncheck all cell line boxes except H1-hESC.**	
10	**Click the "Submit" button at the top** to return to the genome viewer.	
11	**Note that we can now see** that there is signal associated with this histone mark in stem cells in this region. This was difficult to examine before because of the other color overlays.	
12	**Return to the histone mark page** by clicking the gray bar to the left of the browser track. **Turn on or off various cell lines** to view the data. **Return to the viewer** each time by clicking "Submit".	
13	The various data types in this region should help you to understand possible features of regulation of the genes in this area.	
This exercise was inspired by the Figure 5 illustration in the ENCODE User Guide paper. See that figure legend and the accompanying text for more assessments of the data and the features in this region: **http://bit.ly/plosENC**		