# Exercises for the ENCODE data in the UCSC Genome Browser

**1) Visually examine ENCODE project data for information on possible regulatory aspects of genes in the 3p21.1 region of the human genome. Examine some Transcription Factor Binding Site data, and some RNA binding protein data.**

*Skills: Identify ENCODE data; visualize transcription factor binding site and RNA binding data. Become aware of the different cell types in use.*

**2) Examine a region of the X chromosome for genomic structural variation. Determine if the pattern of structural variation is different between cell types. Explore the genes in this region and consider the implications of variations in this region.**

*Skills: Explore Copy Number Variations (CNVs) in the genome. Consider the different patterns in different cell types.*

**3) Examine the differences in RNA localization data for the TPTE gene. Explore the different segments of the mRNA that appear in the nucleus, the cytoplasm, and in the polysome fractions. Determine if the pattern of exon usage appears to vary in different cell types.**

*Skills: Explore RNA localization data. Consider the different patterns in different cell fractions. Use the filters to set the specific data combinations you wish to visualize.*

**UCSC ENCODE Exercises, version 2.**
**Correspond to the data available in July 2010.**

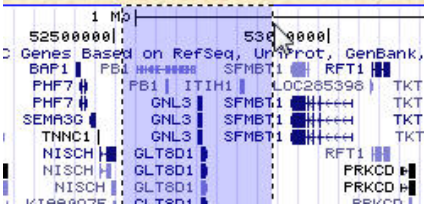


**Seattle**
12600 SE 38th Street, Suite 230
Bellevue, WA 98006
(425) 401-1400

**Boston**
65 Main Street
Somerville, MA 02145
(617) 627-9398

**San Francisco**
193 Haight Street
San Francisco, CA 94012
(415) 252-1519

**1) Visually examine ENCODE project data for information on possible regulatory aspects of genes in the 3p21.1 region of the human genome. Examine some Transcription Factor Binding Site data, and some RNA binding protein data.**

| 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 **March 2006** assembly. Enter the text **3p21.1** in the position box. Click **submit**. | |
| 5 | Get a sense of the scale in this region. It contains many genes, which could presumably be regulated in a variety of ways. | |
| 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.)* | |
| 7 | Add back 3 tracks to the viewer:<br>• **UCSC Genes in pack** visibility (from the Genes and Gene Predictions group)<br>• **Yale TFBS in dense** (from the Regulation group)<br>• **SUNY RBP in dense** (from the Regulation group)<br>**Click a refresh button** to add these tracks back to the viewer. It may take a while for this to load, as there is a lot of data here. | |
| 8 | Let's now scan this region for information on the possible regulation features. Explore the Transcription Factor Binding Site area at the bottom of the graphic. Can you **locate some genes that appear to have potential binding sites for c-Fos? For JunD? Gata-1?** If you don't see a transcription factor data set, open the track details page to see if it is available. You may need to check the box for that factor, click submit, and examine the display again. | |
| 9 | Explore the RNA Binding proteins area. Can you detect binding of different RNA binding protein in different regions? | |
| 10 | **Zoom in** to a region that appears to have Yale TFBS signal data and SUNY RBP ELAVL1 data. I'll click my mouse in the reference sequence track around the beginning of the 1Mb scale bar at the top, and drag for about an inch. It will resemble this:  <br>Release the mouse to zoom in to that region. Look for genes with different patterns of possible regulatory aspects. | |

**2) Examine a region of the X chromosome for genomic structural variation. Determine if the pattern of structural variation is different between cell types. Explore the genes in this region and consider the implications of variations in this region.**

| Step | Action | ✓ |
|---|---|---|
| 1 | **Return to the Gateway page** by clicking the Genomes link at the top of the browser page. You should still be using the Human genome data. | |
| 2 | **Use the Click here to reset** link to return the browser to the default settings. | |
| 3 | **Choose the March 2006** assembly. **Type this region** in the position box: **chrX:45,908,749-80,283,771** | |
| 4 | **Click submit.** | |
| 5 | **Click the "hide all" button** to begin to focus on only the data we are interested in right now, and remove unnecessary data from the view. | |
| 6 | In the Variation and Repeats track group, **click the hyperlink for Common Cell CNV** to access the track details page. | |
| 7 | Here we will change 3 settings.<br>At the top, set **Maximum display mode to Pack.**<br>Set the **Regions menu to Pack.**<br>Set the **Signal menu to Pack.** | |
| 8 | Note that there are 3 cell lines selected. Leave these settings as default. **Click one of the Submit buttons.** | |
| 9 | **Explore this area.** Identify the regions that appear to have heterozygous structural variations. Find potential homozygous deletions. Find potential amplification regions. Note that there are differences between the cell types that are visible here. | |
| 10 | Zoom in to the region that contains the 2 possible amplification areas in the HepG2 cells. | |
| 11 | Let's add the Gencode genes track to the view to examine the genes in this region. In the Genes and Gene Predictions track group, find Gencode Genes. **Select Pack from the Gencode Genes menu, then click refresh.** | |
| 12 | Consider the implications of structural variation in this region and for some of these genes. | |

**3) Examine the differences in RNA localization data for the TPTE gene. Explore the different segments of the mRNA that appear in the nucleus, the cytoplasm, and in the polysome fractions. Determine if the pattern of exon usage appears to vary in different cell types.**

| Step | Action | ✓ |
|---|---|---|
| 1 | Return to the **human Gateway search page**, using the Human **Mar. 2006** assembly. **Click the reset link** to return all settings to default. | |
| 2 | **Search for TPTE using the gene box.** Select TPTE from the gene list. **Click submit.** | |
| 3 | **Click the "hide all" button** to clear the view, and then we'll add back the items we want to see. | |
| 4 | In the **Expression track group**, locate the menu for the ENCODE **Affy RNA Loc track. Choose pack from the menu. Click refresh.** | |
| 5 | Explore the different patterns of RNA signals that have been detected in different fractions. Note that you can examine total RNA, poly-A RNA, and you can find nuclear, cytosolic and polysome fractions, as well as in whole cell samples. | |
| 6 | Let's add back the TPTE gene to examine the intron / exon structure of this gene. In the Genes and Gene Predictions group, **find the Gencode genes track**. **Select full** from the menu. **Click refresh.** | |
| 7 | **Identify the total RNA in GM12878 cells track.**<br>**Identify the total RNA in K562 cells track.**<br>Let's change the view to include only those tracks so that they are easy to compare side-by-side. | |
| 8 | **Click on the hyperlink for the Affy RNA Loc track** in the Expression track area. This will load the details page for this track where we can choose which specific items to display. | |
| 9 | From the list (not in the table above the list), leave GM12878 Cell, Total RNA boxes selected (the top 2).<br>**Uncheck all other GM12878 boxes.** | |
| 10 | Leave the K562 Cell, Total RNA boxes selected (the first 2 in the K562 section).<br>**Uncheck all other K562 boxes.** | |
| 11 | You should have 4 boxes selected at the end of this. **Click the Submit button.** | |
| 12 | On the Genome Browser, see if you can detect possible variations in exon usage in these two different cell types. Also see if you can find exons in the Gencode transcripts that don't appear in these cell types. | |