

Exercises for the UCSC Genome Browser Introduction

1) Find out if the mouse *Brca1* gene has non-synonymous SNPs, color them blue, and get external data about a codon-changing SNP.

Skills: basic text search; Genome Viewer pulldown menus; filters; links to external resources

2) Find the protein sequence for rat leptin. BLAT this sequence vs. the human genome to find the human homolog. Look for SNPs in the coding region of this gene—are there any? Obtain the human DNA sequence for this region, and underline the SNPs.

Skills: obtaining protein sequence; BLAT; finding SNPs in exons; “get DNA” sequence with extended case/color options

3) Find the genomic region for the human NRAS [neuroblastoma RAS viral (v-ras) oncogene homolog] gene. Add 1000 bases to each end of the position in the window. Turn on a Transcription Factor Binding Site (TFBS Conserved) track and look for possible binding sites in the promoter region. Determine if structural variation has been indicated in this genomic region by visualizing the copy number variation (CNV) data from the DGV database track.

Skills: examining TFBS; Copy Number Variations.

UCSC Exercises, version 16a. Correspond to the data available in April 2009.

**The materials and slides offered are for non-commercial use only. Reproduction, distribution and/or use for commercial purposes are strictly prohibited.
Copyright 2009, OpenHelix, LLC.**

Step-by-Step instructions for the UCSC Genome Browser exercises

1. Find out if the mouse *Brca1* gene has non-synonymous SNPs, color them blue, and get external data about a codon-changing SNP.

Step	Action	✓
1	Go to the UCSC Genome Browser homepage, genome.ucsc.edu	
2	Enter the Gateway, by clicking the Genome Browser link from the homepage.	
3	Select Mammal as the clade, Mouse as the genome. Choose the July 2007 assembly.	
4	Enter the text Brca1 in the text box. Click Submit .	
5	From the results list, click a link that appears to be the real <i>Brca1</i> . I will choose <i>Brca1 uc0071pe.1</i> , breast cancer 1 for this example. Examine the gene structure, and look at the SNP track in the viewer with default settings.	
6	From the Genome Viewer page, scroll to the SNP pulldown menu (way at the bottom of the page; Variation and Repeats group area). Click the <u>SNPs (128)</u> link text above the menu.	
7	On the new track settings page you will see many choices about the appearance and features of SNPs you can display.	
8	In the Color Specification area, select “Function” as what we want to color. Change all the menus for types to “black” except Coding Non-Synonymous, which we will make “blue” . Select SNP Display mode as “pack” . Click Submit (near the top) to make the changes back in the browser.	
9	Examine the SNPs track now . Your display should now show all the SNPs (in Pack mode). You should be able to quickly identify SNPs which are coding region + non-synonymous changes.	
10	Select a blue SNP from the display. For this example I will select the blue SNP in the second column: rs28273098 . Click this SNP .	
11	On the new SNP details page, examine the SNP data. Click the link for dbSNP .	
12	A new window should open with that SNP entry in the dbSNP database . You can learn additional details about this SNP, its source, and more from dbSNP.	
<p>Note: UCSC calls the track <u>Simple</u> Nucleotide Polymorphisms--not just “<i>single</i>”--because this data may include insertions, deletions, and alterations that are larger than one single nucleotide position.</p>		

2) Find the protein sequence for rat leptin. BLAT this sequence vs. the human genome to find the human homolog. Look for SNPs in the coding region of this gene—are there any? Obtain the human DNA sequence for this region, and underline the SNPs.

Step	Action	✓
1	From the Gateway page, search for the rat leptin gene, using the most current rat assembly. Pick the lep gene link from the results.	
2	From the rat leptin Genome Viewer region, get the protein sequence for leptin from the UCSC gene details page. Click the gene icon, and find the “Sequence” table near the top of the page. Click the Protein (167 aa) .	
3	Copy the rat leptin protein sequence. You can take everything on the page—the FASTA formatting is fine. Use the Back button to return to the Known Gene page.	
4	Now access the BLAT tool either from the Known Gene page, or the home page, by clicking BLAT in the top navigation bar area.	
5	Paste your leptin sequence into the BLAT text box	
6	Check the BLAT options, and choose to BLAT against the human genome , using the March 2006 assembly , with a protein sequence. All other settings leave as default.	
7	Submit your BLAT search.	
8	From the BLAT results page, click the DETAILS link for the top hit . Examine the details page to see if the match is good.	
9	If your match is acceptable, return to the BLAT results page. Now click BROWSER to see the Genome Viewer location with this match.	
10	If you are convinced we are in the right genomic region in the viewer, we will get the DNA sequence for this region and find SNPs in exons.	
11	First, choose the HIDE ALL button so we can add back just the items we care about.	
12	Next, choose to see UCSC Genes in Full mode, and SNPs (128) in Pack mode. Click Refresh to enforce these changes. In your viewer there should only be 2 tracks now.	
13	Let’s look at the SNPs in the context of the genomic sequence. Click the link for DNA in the blue navigation bar at the top.	
14	From the Get DNA page, click the Extended Case/Color options button.	
15	Choose BOLD for UCSC Genes, and Underline the SNPs. Also put 255 in one of the color boxes for SNPs.	
16	Submit. You should have a new page with your sequence, with the UCSC Gene exons in bold, and SNP locations underlined and in color.	
Special note: Extended case/color options list only those tracks which are currently shown in the Genome Viewer window.		

- 3) Find the genomic region for the human NRAS [neuroblastoma RAS viral (v-ras) oncogene homolog] gene. Add 1000 bases to each end of the position in the window. Turn on a Transcription Factor Binding Site (TFBS Conserved) track and look for possible binding sites in the promoter region. Determine if structural variation has been indicated in this genomic region by visualizing the copy number variation (CNV) data from the DGV database track.

Step	Action	✓
1	From the Gateway page, search for the human NRAS gene in the March 2006 assembly. Choose the NRAS gene link for uc009wgu.1 from the search. <i>Click the default tracks button to restore default view if needed.</i>	
2	From the NRAS Genome Viewer region page, change the values in the position box to be 1000 nucleotides less on the left or 5' side, and 1000 nucleotides more on the 3' side. That is: Original numbers: chr1:115,048,601-115,061,038 New numbers: chr1:115,047,601-115,062,038 Click the Jump button after making the number changes.	
3	Scroll down to the Regulation track section. Find the TFBS Conserved track menu. Select full from the pulldown menu. Click a refresh button.	
4	Examine the viewer again. Find the location of the new TFBS data. <i>Note: Keep in mind the direction of transcription of this gene as indicated by the arrowheads on the gene track.</i>	
5	Click the top-most TFBS tickmark. This should be V\$EVI1_06. The details page will show the name as Factor: Evi-1.	
6	Return to the main viewer page.	
7	Determine if structural variation has been indicated in this genomic region by visualizing the copy number variation (CNV) data from the DGV database track. In the Variation and Repeats section, turn the DGV Struct Var menu to pack.	
8	Click the top-most variation item (3306) to learn more about it.	
9	On the 3306 page, click the DGV Browser and Report: 3306	
10	At the DGV site, determine the kind of variation this represents. Variation Type: CopyNumber	