**Xiaoxu Na**

**Intro: Basic Genome Browser navigation**

Here we will explore the Genome Browser to get familiar with moving around the landscape and load a simple Custom Track

**Navigation -- find a gene**

There are several ways to get started. We will start with a clean "cart", select a genome assembly and enter a gene name.

**Steps**

1. Go to <http://genome.ucsc.edu> and hit the button "Genomes” at the upper left in the blue bar.
2. Select assembly hg19
3. In the "Position/Search Term" box, type "PIGP". Then hit the “go” button

**Navigation -- zooming, panning and controlling data tracks**

To get around in the browser view, you can move to any continuously variable region of your choosing.

**Steps**

1. Above the browser image, hit the "zoom out 3x" button to see the region around the gene.
2. Click on “default tracks”
3. Data in the various "tracks" are drawn from separate tables in the browser database.
4. Note the spikes of conserved sequence among the species in the Conservation track.
5. Below the image, in the track controls, find these tracks and select "hide": RefSeq Genes, Human mRNAs, Spliced ESTs, Repeat Masker. Select "Refresh" from any blue bar.
6. Click on “PIGP” name to see rich information about the gene.

**Question 1:** Give a brief description about PIGP gene based on information on UCSC genome browser.

Homo sapiens phosphatidylinositol glycan anchor biosynthesis, class P (PIGP), transcript variant 1, mRNA.

This gene encodes an enzyme involved in the first step of glycosylphosphatidylinositol (GPI)-anchor biosynthesis. The GPI-anchor is a glycolipid found on many blood cells that serves to anchor proteins to the cell surface. The encoded protein is a component of the GPI-N-acetylglucosaminyltransferase complex that catalyzes the transfer of N-acetylglucosamine (GlcNAc) from UDP-GlcNAc to phosphatidylinositol (PI). This gene is located in the Down Syndrome critical region on chromosome 21 and is a candidate for the pathogenesis of Down syndrome. This gene has multiple pseudogenes and is a member of the phosphatidylinositol glycan anchor biosynthesis gene family. Alternatively spliced transcript variants encoding different isoforms have been described. [provided by RefSeq, Feb 2016].

**Custom Tracks -- add your own data**

We will load a simple Custom Track to see how user-supplied data can be added to the data already in the browser.

**Steps**

1. Below the browser image, select "add custom tracks."
2. In the upper box, type (or paste):

track name="test custom track" visibility=pack

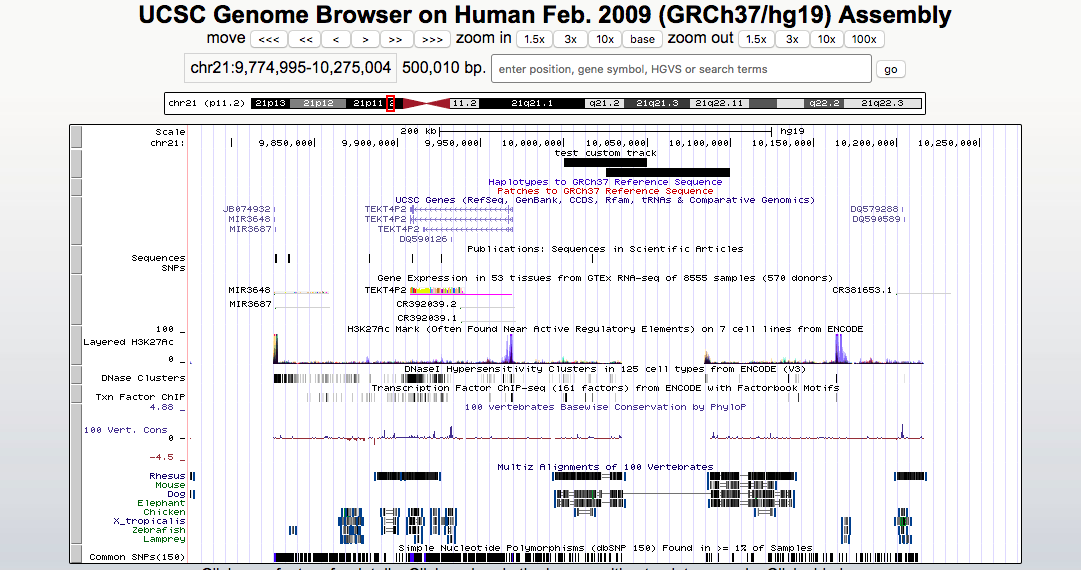
chr21 10000000 10050000

chr21 10025000 10100000

Select "Submit."

1. At the browser, select "zoom out 10x"

**Question 2:** Take a screenshot

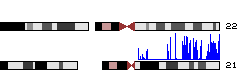


**Tutorial: Viewing sequencing data in the UCSC Genome Browser**

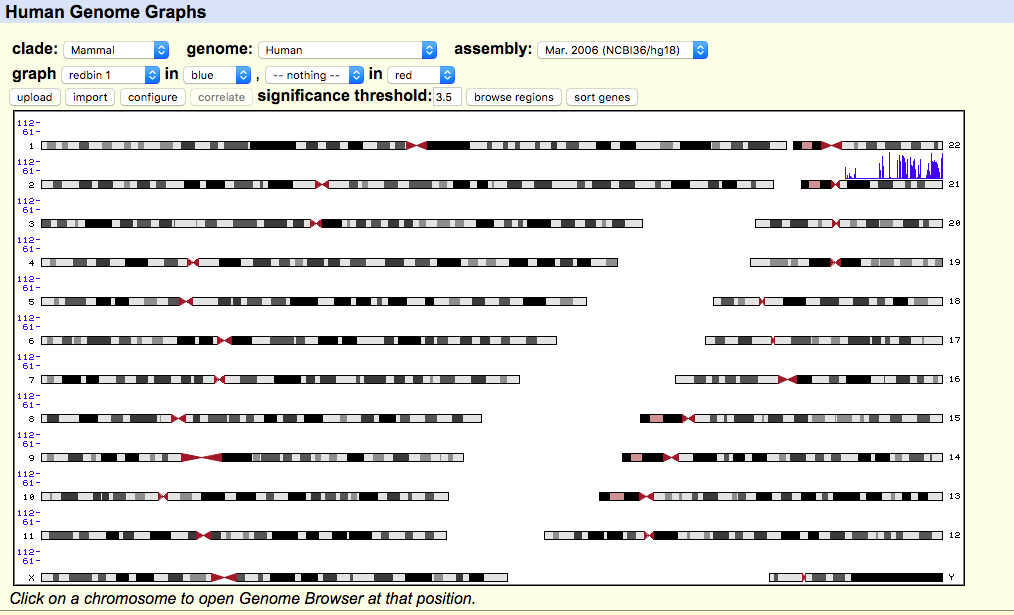
*Genome Graphs with .sgr data .*

Steps

1. In **a new web browser window or tab**, go to http://genome.ucsc.edu/
2. Click on “Tools” and select "Genome Graphs” on the sidebar.
3. In the assembly menu, change the selection to "Mar. 2006 (NCBI36/hg18)".
4. Click the upload button.
5. In the input labeled "name of data set", type or paste "redbin".
6. Download a file called *chr21\_extended.txt\_redbin.sgr* from Blackboard into your computer and upload it to Genome Graphs by clicking on Choose File button. Then click submit. You should see confirmation screen; click OK to return to the Genome Graphs display.
7. Now you should see a layout of human chromosomes. Genome Graphs doesn't show the data until you select it. In the graph menu on the left, change "--nothing--" to "redbin 1". Now the data appear -- in this case, for chr21 only.

[](http://genome.ucsc.edu/cgi-bin/hgGenome?hgS_doOtherUser=submit&hgS_otherUserName=AngieHinrichs&hgS_otherUserSessionName=ABRF2010_GG)

**Question 3:** Take a screenshot of your display



Now we will use Genome Graphs to navigate to the Genome Browser.

1. Click on a peak to jump to the Genome Browser, displaying a 1,000,000-base region of the genome corresponding to the location you clicked in Genome Graphs.
2. Click on the top label "User Supplied Track 1" to make the peaks taller and more visible.
3. In general, peaks correspond to upstream regions of genes, which make sense for RNA polymerase II ChIP-seq. (The *chr21\_extended.txt\_redbin.sgr contain the results* polymerase II ChIP-seq)

**Question 4:** Take a screenshot of your display and list at least one gene that has RNA polymerase II peak.



c21 or f59

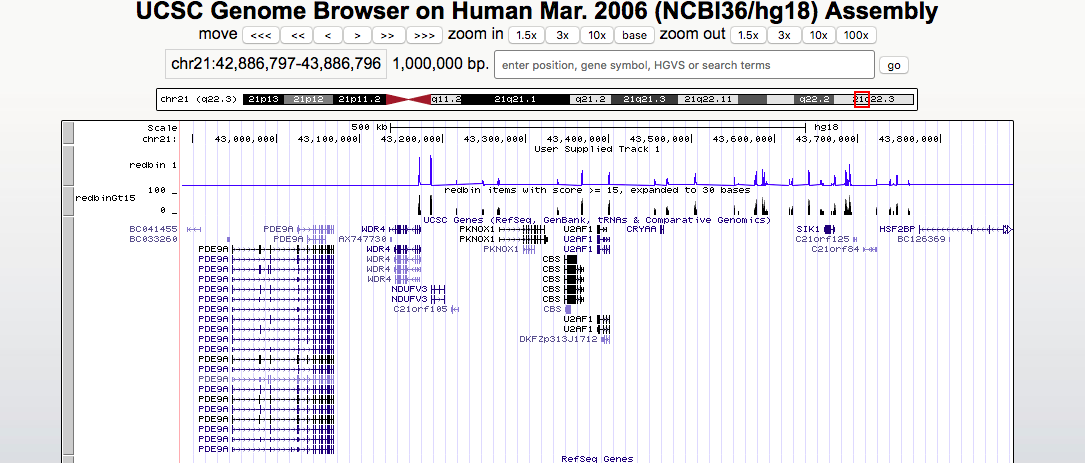
**Transforming .sgr into bedGraph for Genome Browser viewing**

The Genome Browser can draw a somewhat nicer display of .sgr data, if we do a simple transform of the .sgr data into a similar format called bedGraph. bedGraph has both start and end genomic coordinates, so scores can be drawn over regions as opposed to single bases.

**Steps**

1. In the Genome Browser, click the "manage custom tracks" button below the image.
2. Click on "add custom tracks"
3. Download data *redbinGt15.bed* from Blackboard and upload it by clicking on Choose File button. Then click submit
4. Click on “go to genome browser”
5. Now the new custom track shows grayscale representations of the score. You can click on the track title ("redbin items with score > 15") to expand the display to bar graph, which should look a lot like the "redbin 1" track created in Genome Graphs.

**Question 5:** Take a screenshot of your display.



**Viewing alignments in BAM format**

**Steps**

1. In the Genome Browser, click the "manage custom tracks" buttom below the image.
2. Click the "add custom tracks" button.
3. Paste this text into the box:

**track name=chr21\_export type=bam bigDataUrl=http://genome-test.cse.ucsc.edu/ABRF2010/chr21.bam visibility=dense**

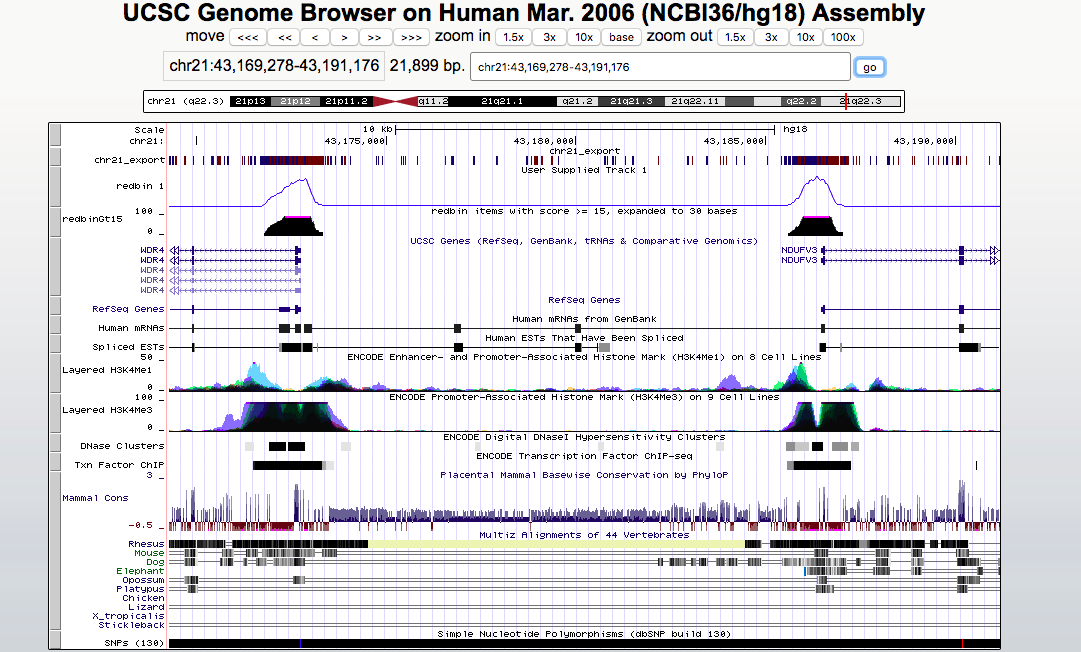
1. Click submit. Click "Genome Browser" in the blue top bar. Now you should see an almost-solid dark blue bar with a few dark red stripes, labeled "chr21\_export" above and to the left.

[catch up](http://genome.ucsc.edu/cgi-bin/hgTracks?hgS_doOtherUser=submit&hgS_otherUserName=AngieHinrichs&hgS_otherUserSessionName=ABRF2010_chr21_export_1MB)

There are so many alignments in this 1MB window that the items can't be drawn individually -- the image would become too tall. Let's view a smaller genomic region to get a better look at the alignments.

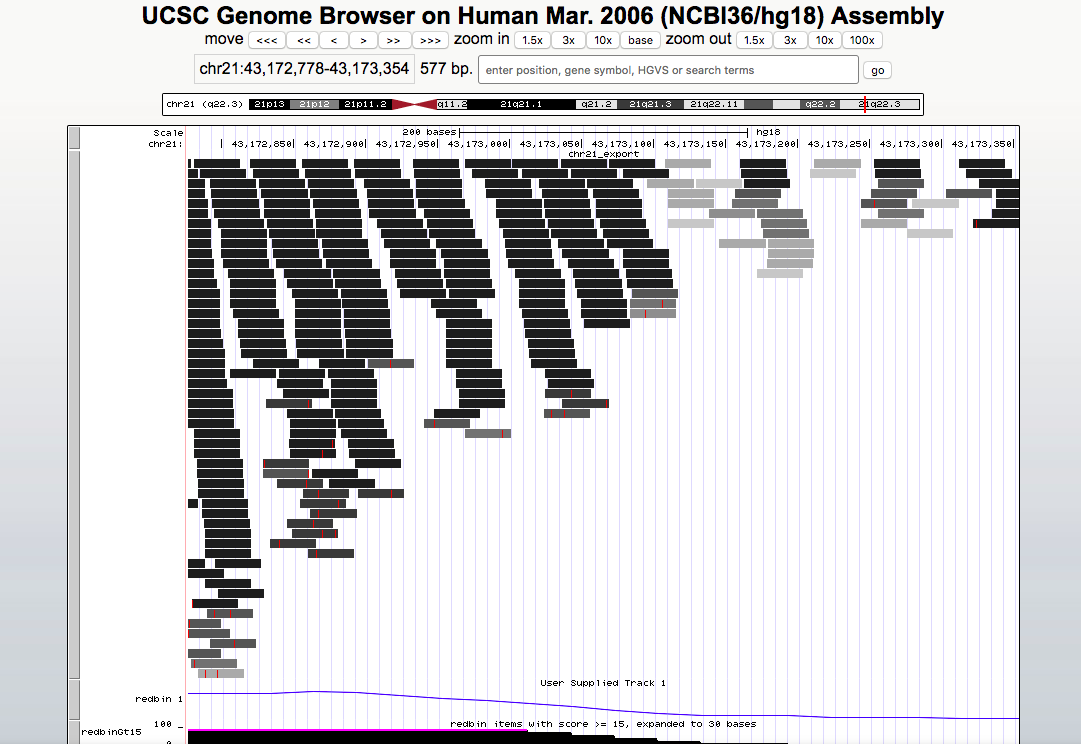
1. Paste **chr21:43,169,278-43,191,176** into the position box and click "go".
2. Now we are viewing about 22,000 bases covering two peaks, and there are still so many alignments that the track is forced into "squish" display mode: items are shown at half-height and read names are suppressed.

**Question 6.** Take a screenshot of your display.



1. Let's zoom in again, jump to position **chr21:43,172,778-43,173,354***.*
2. Mismatches appear as light red tickmarks. They are a bit easier to see if the items are drawn in grayscale, so again click on the light grey bar on the left to return to track controls.
3. Under "additional coloring modes", select "Use gray for [alignment quality]", and click submit. Then change display mode to “pack”.
4. Note the trend for mismatches to appear more often in alignments with lower quality, and also the relative frequency of lower-quality reads at the edge of the peak.

**Question 7:** Take a screenshot of your display. You should see the trend for mismatches to appear more often in alignments with lower quality, and also the relative frequency of lower-quality reads at the edge of the peak

****

**Finding variants in alignment display**

In some regions, stacks of red tickmarks (mismatches) indicate HeLa variants with respect to the reference genome assembly. Here are several examples.

1. Paste **chr21:37,366,724-37,366,992** into the position box and click "go".
2. To compare with variants in dbSNP, scroll down to the Variation section and change the visibility of SNPs (130) to “pack”. Click "refresh" on any blue section header.
3. The SNP track appears at the bottom of the image. SNPs classified "coding non-synonymous" by dbSNP are colored red.
4. To see a variant not in dbSNP, paste **chr21:37,367,831-37,367,845**.

**Question 8:** Take a screenshot of SNPs (130) in the region and the variant not in dbSNP.



Reference: Base on ABRF2010 Tutorial