Multi-Region mode: Viewing discontinuous regions in the UCSC Genome Browser

This video will demonstrate the Multi-Region display mode on the UCSC Genome Browser.

In particular, we will show the feature that allows display of multiple, discontinuous regions in the genome together on the Genome Browser. We will work with the human assembly hg19 because it has some nice RNA-seq data to view, but the method is generally applicable to any genome assembly. We will use a set of genes as the source of the coordinates for the regions to view. This feature will allow you, for example, to look at your RNA-seq data for a pathway or other group of interesting genes, all at the same time, even though they are not together in the genome.

Go to the Browser main page, genome.ucsc.edu.

[0:41 Set the Genome Browser to defaults]

Click the Genome Browser link to go the Gateway page, which allows access to all the organisms hosted on the Browser.

Reset the Browser to start at the default position using the "Genome Browser" pulldown in the top menu and "`Reset All User Settings`."

[1:00 Go to the Genome Browser at hg19]

Select "`hg19`" in the "Human Assembly" pulldown menu on the right half of the page. Hit the "`Go`" button. At the Genome Browser graphical viewer, clean up the image using the "`hide all`" button below the Browser graphic.

[1:21 Turn on gene and RNA-seq tracks]

Now let's turn on the gene track, "UCSC Genes" to "`pack`" and hit the refresh button. We will now find an RNA-seq track to make our display more interesting later. Below the Browser graphic, find the "Expression" bluebar. In alphabetical order, on the right side of the screen, find the track "ENCODE RNA-seq…." Click the underlined link to access the configuration page for this track. On the next page, click the link, "`Cold Spring Harbor Labs Long RNA-seq`."

On this page, set "Maximum display mode:" to "`full`." Turn off all the cell lines first, by clicking the minus sign next to the word "All" in the upper left corner of the box. Next, select these three cell lines: GM12878, H1-hESC and A549.

Below the box, find the pulldown menu labeled "RNA extract:" and select "`PolyA+`", then "`close`." In the list of tracks below, unclick each of the three tracks with Replicate rank, Pooled. You should now have a dozen tracks in the list. Change the visibility for each to "`full`."

Now finally, hit "`Submit`." In the Browser graphic, we see the three cell lines, each in a different color, with replicate data on each strand, for a total of four data tracks per cell line. Most of the expression signal for the SOD1 gene is in the pair of tracks representing alignment to the plus strand.

[3:13 Get gene coordinates using the Table Browser]

Let's go now to the Table Browser using the "Tools…" menu in the top bluebar. We will use a list of genes as raw material, and extract the coordinates for those genes and use those coordinates to limit our view on the Genome Browser.

The Table Browser defaults to the UCSC Genes set on hg19 (on hg38 it is GENCODE v24). The default table is preselected: knownGene. This table could be used, but it will give us a set of coordinates for each isoform, which we'd then have to trim down to a single isoform per gene. Instead, we can use the knownCanonical table to extract a single isoform and a single set of coordinates for each gene. This is the table which underlies the function in the main Browser display that allows you to turn off multiple isoforms for each gene.

Be sure the "genome" button is selected (which it will be if you have just reset the defaults as we have, but it may not be in other circumstances). If you query a narrow region with a list of genes, it will not work properly if the genome button is not selected, thus returning results only for genes in that small region.

[4:27 Paste list of genes into Table Browser]

Now let's use the "paste list" button to load in our list of genes.

We will paste in a list that includes one FGF factor (FGF2) and several of the FGF receptor genes (FGFR 1 through 4), and submit.

[4:55 Extract coordinates of genes]

To get just the coordinates of the genes, we will want to extract selected data fields from the table, so using the "output format" pulldown menu, choose "selected fields from primary and related tables." Get output.

On the next page, select the boxes for "chrom", "chromStart" and "chromEnd" from the knownCanonical table and "geneSymbol" from the kgXref table. Get output. Now we have the coordinates of our list of genes.

[5:31 Reorder gene list and add header lines]

Let's copy the information to a text file, then use the Back button to return to the Table Browser. We will reorder the text list of gene coordinates into the order we prefer. The display in the Browser will follow the order in the list.

The next step is to add some header information to tell the Browser how to behave when we enable Multi-Region mode.

[6:04 Get help for Multi-region mode]

You can find this information yourself in the help docs as follows: Go to "Help" in the top bluebar menu and select "FAQs and search."

Go to the text box next to “Search the entire Genome Browser website:” and type in “multi-region” and submit. Click into the top link, Genome Browser Multi-Region. This is a comprehensive guide to using this feature.

On the new page, under “About the Multi-Region modes,” select “Custom Regions.”

Here we see that there are three types of header information that we can use at the top of our BED file: database, short description and padding.

We will type in at the top of our text file of gene coordinates the following:

```
#database hg19
#shortDesc Group of genes
#padding 1000
```

each with a hashtag in front of it. The padding puts a small amount of space between the regions to keep the annotations from running together in the display. Because these genes are rather large, we use 1000 bases.

Use the link in the top bluebar, Genome Browser, to return to the Browser graphic.

[7:24 Load coordinates into Multi-region]

Going to the top bluebar on this page, select “View… Multi-Region.” The “View” pulldown has options that are sequence-specific, so it is only available on the Browser graphic page. Paste the coordinates with the headers into the text box labeled “enter custom regions as BED....” Select the radio button to enable the feature and check the “Highlight” box at the bottom before submitting.

The error message explains that the new regions are not near where you were before. Turn off the message using the “ok” button. Zoom out by 100x twice to view all the regions. The display stops zooming when all the regions are in view.

Note that we now have our five regions in view, along with all their annotations. Alternating regions have a light blue background because we checked the “highlight” box on the Multi-Region menu. You can see that some of the genes are expressed in some of the cell lines, but not in others. Some of our five genes are transcribed from the plus strand, and some from the minus strand.

Exit Multi-Region mode via the View pulldown menu, or use a keyboard shortcut. The shortcuts are found by typing a question mark (?) on the graphic page. We will see that the default view is attained by typing “dv.” The Browser will go to the first region in our list and we are now able to navigate anywhere we wish in the genome.

Thanks for watching and thanks for using the UCSC Genome Browser.