Controlling visibility of data tracks on the UCSC Genome Browser

In this tutorial, we will demonstrate some of the most important visibility controls and track configurations in the UCSC Genome Browser. We will begin at genome.ucsc.edu and visit the Browser using the Genome Browser link. At the resulting gateway page, let's "`reset all user settings`", which gives us the human genome assembly, hg38.

[0:32 Set Browser to hg19 UCSC Genes track.]

We will switch to the previous assembly, hg19, and hit the "`Go`" button. We start with a large number of tracks of various kinds turned on. Let's hide them all to begin with, and reset the browser to display just the UCSC genes track. You'll notice that there are five different display options, and we'll begin by discussing the various options using the UCSC Genes track.

Set the UCSC Genes track to "`pack`" and hit the refresh button and we see that we have a single gene on the screen in the default location, and it's the SOD1 gene. Let's go to a different gene location, which has multiple isoforms, the CYP2D6 gene, and we'll see how the visibility settings affect the display under various scenarios. You can see here that there are three isoforms of this gene and an antisense transcript from another gene from a nearby location. We're at pack visibility.

[1:37 BED track: Pack and dense modes.]

We want to look at dense visibility, but before we do, let's take a look over here at this one exon that is differentially expressed. When we switch the gene track to "`dense`", you see that any location that is expressed in any one of the isoforms is covered by an annotation and all exons are collapsed into a single row on the track, regardless of how many different isoforms may be expressing the exon.

Let's zoom out by a factor of 10 and pick up some nearby genes and you can see that in dense mode there are no labels anymore and everything is essentially squeezed into a single row. We are displaying only the footprint of the genes on the genome, without any detail. Now, if we go

back to the pack visibility, we can see that the multiple isoforms of the CYP2D6 gene are shown, as are multiple isoforms of the nearby genes.

[2:37 BED track in full mode.]

Let's switch the visibility now to `full`, and we can see that in `full` each individual isoform of each gene gets a separate row, and the genes are no longer packed tightly into the browser display. Full display mode takes up much more vertical screen real estate than pack.

[2:58 Squish mode.]

The one remaining type here in a typical gene track is squish, which essentially squeezes down the pack display into a much tighter display, dropping all of the item names, and showing you just the individual isoforms in a very condensed fashion.

Let's leave the gene track on squish, and at this point we can point out that the data in this and many other tracks in the Genome Browser are in a format known as BED or the Browser Extensible Display, which is essentially a way of describing locations for putting boxes on the Browser, the simplest being BED3 with chrom, chromStart and chromEnd as the first three columns in the specification for the track.

[3:44 SNPs are usually better in dense.]

Let's turn on now the All SNPs track and we'll turn on All SNPs (147). We'll set it also to `pack`, and you'll see that at this location we have many, many SNPs and they are so tightly squeezed into the track that it's not even possible to display them in `pack` mode. The Browser reverts to squish. So, let's "`zoom in`" now to the region around the original gene, CYP2D6, and you'll notice that in pack mode, the SNP track has a number of tick marks all squeezed together with their RS numbers, their "reference SNP" annotation names.

Let's switch this pack to full now and we'll see "too many items to display" so let's "`zoom in`" farther to a single exon, and you'll see that in full mode now we have this similar display that we had in the UCSC Genes track and that each SNP is on a single line with the labels down

the left side and a lot of white space on the screen. Now for a track such as the SNP track, it is quite common to prefer to not see the actual names of the SNPs but rather collapse it into "dense" mode so that the individual variants are squeezed into a smaller space where you just see the footprint of the track. Let's turn the gene track on to pack and you'll notice that the colored SNPs are all within the same region as the exon. This is because the coloring green or red in a SNP track indicates coding synonymous or non-synonymous, respectively. Let's zoom into another region here even farther, and we'll see that the individual variants are one or two bases in extent and in dense mode we've thrown away all the labels and we just see the footprint of the track.

The final version of visibility on the SNP track is to switch it to squish mode, which, as it did for the genes track, gives us a display that uses up a very small amount of space on the browser.

[5:46 Some tracks have useful mouseover in pack and full modes.]

For some tracks it is useful, however, to have the labels shown and one such track is the "OMIM Alleles" track.

Let's put the SNP track onto "hide" here and we'll move out by a factor of 10 and show you that the OMIM Allelic Variants gives us a display where it's possible to use the mouseover feature of the browser to give us some individual information about the variant, in this case that it's an arginine to cysteine at amino acid 296 and it has an rs number and a specific disease state associated with it, specifically, "debrisoquine, ultrarapid metabolism of."

[6:30 Super tracks and wiggle tracks.]

I'll turn off the Allelic Variants track and move on to discuss another type of track, known as a wiggle track. Let's navigate over to the FGF13 gene and we'll turn on an appropriate track for looking at a wiggle track. A wiggle track is a track that gives us a second dimension in the data. A good wiggle track to use for this on the hg19 genome assembly is in the expression group and is the ENCODE RNA-seq track. You'll notice that the pulldown menu here has only two options, hide and show, and that

is in part because it's a track of a type we call a super track– indicated by a dot dot dot (or ellipsis) on the short label of the track.

This indicates that it's a package of multiple composite tracks. This type of track has three levels of control: at the top, it's a supertrack; in the middle are composite tracks, which consist of multiple individual data tracks. To illustrate what I mean by this, let's click into a link and go to the configuration page, and you'll see that the track is set to hide, but there are five different composite tracks within this supertrack. We'll turn all of them off, except we'll turn the Cold Spring Harbor Lab Long RNA seq track back on to `full`. You'll notice that the display mode of the super track is switched to `show` when we change the Long RNA track to full, but before we display it, let's click into this particular composite track and see all of the individual tracks that make up this composite.

Let's turn them all off by hitting the minus sign and then we'll turn on the H1-hESC, the embryonic stem cell line, and also the K562 line. Both of these were in the ENCODE project, were in Tier 1, the ones with the most extensive data. We'll scroll down below the selection menu and choose just PolyA+ data tracks and we will deselect the Contigs tracks. So, we'll be just looking at plus strand and minus strand wiggle tracks in duplicate.

[8:49 Wiggle tracks in dense and full.]

There will be two replicates for each of the two cell lines. Remember that we set the composite track to full, but we have set the individual tracks to dense.

Now when we hit "`submit`", you'll see that we are displayed in dense mode. You can see how dense mode shows you by varying levels of color, dark versus light, whether you have signals at specific locations or not. The Browser will display the data at the most restrictive display mode, so setting the Composite to full and its component individual tracks to dense will result in display in dense mode.

Let's use the right mouse button and turn on one of these tracks to full and see the data as a second dimension graphic and zoom into the 5'

end of this one isoform shown in the gene track above. Notice that the direction of transcription is right to left for this gene. So, you can see here that when the one track is in full, you've got a very strong signal corresponding to each of the exons shown as a peak in the second dimension and you can compare that to the replicate below it, which is still in dense. Let's take this other track, put it also to `full`, and we see a nice concordance between the replicates.

So let's take a look at what pack mode looks like for these wiggle tracks. You can see that it's essentially the same as full mode except that it takes up slightly less screen in the vertical dimension. Notice that in both pack and full, the little pink hats on the peaks indicate that there is data beyond the maximum vertical setting.

Let's switch these back to `dense` so that we have all four sub-tracks set to dense and try one of them on `squish` mode. You can see that it's essentially the same as dense, but takes up even less vertical real estate. Let's hide all eight of these tracks using the right mouse button on the configuration bar at the left side of the screen and hide the entire track set.

[10:54 Conservation tracks.]

Now let's have a look at yet another type of composite track called Conservation track down in the Comparative Genomics section. We'll switch it to `pack` mode to start with, and hit "`refresh`". For this particular gene we see that we have signal for the multiz alignments of 100 species that align pretty well with the exons and we also have a little graphic up here, which shows the conservation score from the phyloP algorithm.

PhyloP is also in pack mode, but we can switch it from pack to `full`, and you see that it behaves the same way as did the RNA-seq wiggle track earlier, where pack is a shorter version in vertical space than full but essentially shows the same type of graphic.

Using the control button on the left side of the track now, we'll go into the composite track and look into the various sub-tracks. You'll see that the maximum display mode is set at full, which is an overarching control

over the other ones. If this were set to pack, for example, none of the other tracks would actually display at full even if they were set to full.

So, we'll leave the Maximum display mode set at pack, set PhyloP at `full`, and we'll show how the maximum display mode overrides the individual sub-track display mode. While we're in here, let's select all of these 100 species. We'll use the scroll bar here to look at the entire list and notice that all of the species' checkboxes on the left are now checked. You can see the phylogenetic tree on the right side, standard for the Conservation tracks on all genome assemblies.

Now if we hit "`submit`", we'll see that even though the phyloP track was set to full, it is displayed as pack because the overarching composite track was set to pack, which limits the visibility of any of the subtracks that are displayed. We can override the maximum mode at the graphic by setting PhyloP to `full`.

Now we have all 100 species set to full on the multiz track and you can see an interesting pattern here where the introns have less and less homology as you go back through evolutionary time, down the page. Once you get back past the placental mammals the introns drop off very dramatically, but the coding exons maintain their conservation. Note that the long 5'-untranslated region of this exon is also not conserved and in some species the exon seems to be missing entirely.

Let's take a look at what the multiz track looks like when it's changed from `pack` to `full` and you can see that instead of just a coverage track you now actually have individual graphic wiggle-type tracks that show for each species the sequence conservation score.

The other two modes, dense and squish, are not too useful in the multiz as it simply squeezes all of the tracks down into a much smaller space and you lose all of the resolution with respect to the 100 species. So, for this track type, the recommended view is pack, and you can see where the various genomes have coverage.

That concludes our conversation about the various types of visibility modes and how several different track types are affected by the visibility settings.

Thank you for watching and thanks for being a user of the UCSC Genome Browser.