Viewing alternate haplotypes with 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 alternate haplotypes for the hg38 human genome assembly. This method is generally applicable to other assemblies with alternate sequences.

[0:17] Go to the main page, genome.ucsc.edu. 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.

[0:34] To see the alternate sequences available in the official NCBI reference assembly for hg38, click the "`view sequences`" link on the gateway page.

Scroll down the page below the complete chromosomes and check out all the alternate sequences that have become part of the reference assembly. This is part of the effort by the Genome Reference Consortium to give the reference better representation across all populations. These alternate sequences can be viewed directly in the Browser, where they are treated as separate chromosomes, but there they lose their chromosomal context.

There are many alt sequences for several chromosomes. Let's take the first one on chr1, chr1_GL383518v1_alt, which is 182 kb in size. Let's click into the alt and look at it in the Browser. Notice that there are several genes annotated in the region, including INTS3 and GATAD2B. Hit the "`hide all`" button to turn off all the tracks. Let's resize the screen to make the most of the width of the web browser.

[1:44] Most of the alts have some sequence on the ends that allow us to locate them on the appropriate reference chromosome. We will download 25 kb from each end and Blat it back to the genome. That will give us an anchor so that we can see the effect of the multi-region function. Click the coordinate box on the left side of the section just above the chromosome ideogram. The coordinates now appear in the text box on the right. Let's save the name of this alt to a text file for later use.

We will pull 25 kb from the right side of the alt first. Edit the sizes by changing the start coordinate which is now 1, to 157 thousand by subtracting 25 thousand from 182 k. We confirm that the size is 25 kb and then go to "`View… DNA`" in the blue bar at the top of the page. "`Get DNA`" displays an ASCII FASTA file which we will copy/paste into a text editor. We will edit the name of the sequence by adding "_rt" to the end of it. Now we will get 25 kb from the left side of the alt and copy/paste it into the text editor.

[2:55] Going back to the Get DNA page, we will edit the coordinates to: "1-25,000" and "`get DNA`". Copy, paste, and once again edit, this time adding "_lf", indicating it's the left side.

[3:18] Now we'll copy both blocks. Use the Back button to get back to the Get DNA page. In the "`Tools…`" menu, "`BLAT`" we paste everything. Two times 25 is the maximum for web-based Blat. Now "`submit`" and Blat finds matches on the alt chromosome as well as on the full-length chr1. You'll see that it's a perfect 25,000 base match on the alt chromosome as you would expect, and it's a 17,000 base match on chr1. This match is a 99.9% identity.

Further down the page we see that the right end has a perfect match on both chromosome1_alt and the main chr1. Let's click into the match on chr1 and navigate to the Browser.

You see that this alignment is at 153 mb on the right arm of chr1 and it's a perfect match. If we zoom out by 10x we'll have a 250 kb window, which we can drag over to see both ends of the Blat match at once. You can see that the matches on the left show some overlap, that there's a small amount of repeat in this region, and that the red tick marks indicate that there is some mismatch between our original sequence on the alt and this location on chr1.

[4:37] Now we are well positioned to see exactly how the alt is substituted into the genome and we will now turn on the GENCODE Genes track to observe the genes in the region. Notice that once again some of the same genes appear: the GATAD2B gene and the INTS3 gene.

Going to "View… Multi-Region," select "Show one alternate haplotype, on its chromosome…" and paste in the name of the alt we saved earlier. And "submit." The new window is 547 kb, three times the size of the alt sequence in the middle. Note the light pink lines marking the inserted sequence. Note also that the alt chromosome comes with all its own annotations.

The coordinates in the Base Position region at the top of the image reflect the coordinates of the alt region in the middle, while the coordinates of the whole chromosome on either side reflect those coordinates. Click the coordinates of the virtual chromosome above the Browser graphic to see the database representation of the regions in the image: two from chr1 and one from the alternate chromosome.

Note also the perfect alignment with no red tick marks in our Blat alignment on the left side. This is because the sequence from the original alt that we used is a perfect match to the alt that is now substituted into the chromosome. We do not usually recommend using the back button when a suitable alternative is available, but we'll return to the larger region using the back button.

[6:10] Some tracks will not have annotations in the alt region. For example, turn on the ENCODE Regulation track, in the Regulation blue bar group below the Browser graphic. This track was lifted from the hg19 genome assembly and you'll notice that the track has no annotations in the alt region.

[6:31] To remove the alt sequence from the image, return to "View… Multi-Region" and select the button at the top, next to "Exit multi-region mode" and "submit". Note the left side alignment is full of red tick marks again, indicating the mismatch between our probe from the alt and the main reference chr1. You'll also note that the annotations from the ENCODE Regulation track are present on the main chr1 in the Browser graphic.

Thanks for watching and thanks for being a UCSC Genome Browser user.