How do I find the SNPs upstream from genes using the UCSC Genome Browser?

This tutorial will show how to find all the single nucleotide polymorphisms upstream from genes using the UCSC Genome Browser.

The Genome Browser is found at genome.ucsc.edu . This is the main index page in the upper left-hand corner of which are two ways to get to the main Browser graphic.

"`Click here to reset`." We'll set the Browser to the default location and the default sets of data. Using the [`submit`] button on the right-hand edge of the screen, we'll go into the main Browser graphic which shows a large number of data tracks.

[0:43] We'll start by hiding all of the data tracks, which simplifies the view and then we will turn on the UCSC Genes track to "`pack`" and that will show us a single gene at the default location, the SOD1 gene. So let's navigate to a more interesting location, the NBAS gene, and we will type that in and then select it from the offered menu. And the [`go`] button then takes us to that location in the genome. The reason why this is more interesting is that it has a number of different start sites and we are going to make a custom track out of the upstream regions of the genes and then use that custom track to intersect with the SNP table from dbSnp so that we can find all the SNPs in those regions. And it is also interesting because to the right-hand side of this gene is another gene, the DDX1 gene and the two of them are transcribed in opposite directions so it's a nice easy location to observe that we really do have the upstream regions before we proceed with the intersection. Let's zoom out by a factor of 10 now and we will get a number of other genes onto the screen and we will use this region for our inquiry using the Table Browser.

[2:00] Under the "`Tools…`" pull-down menu we can navigate to the Table Browser and at this place we have the default track, "`Genes and Gene Predictions`" group and UCSC Genes track, turned on and we will use the [`position`] button to select the region where we were just viewing over on the Genome Browser graphic. The Table Browser remembers where you were on the Browser graphic. So to get the custom track for the upstream regions we can go directly to a custom track if we'd like but for this demonstration we will instead go to "`browser extensible data`" as an intermediate step. We will be able to see our data on the screen before we go to the custom track.

So the BED file output is selected and then the [`get output`] button takes us to a choice page where we can set some parameters. We'll create one BED record for the upstream region and we'll accept the 200-base upstream default position but of course you're free to use whatever region you want. We will include a custom track header which will make it easy to copy into the custom track input page and we will name the track "`upstream 200 bases`" and we will call the long label "`200 bases upstream from genes`." And then the [`get BED`] button will take us to a page where we can see the data. So each one of the transcripts has an individual item that's 200 bases in extent on this page. So we will "select all" on this page and then copy all of this back into the Genome Browser. The [`Back`] button takes us back to the Table Browser and then the

Genome Browser link takes us back to the Browser graphic.

[3:38] Below the Browser graphic is an [add custom tracks] button which allows us to paste this content into the box and you can see that we have the header created from our choices back at the interface. The [submit] button gives us a custom track and then clicking on the position link takes us to the first item in the custom track. You can see that it's 200 bases and we will zoom out by a factor of 100 and we will see that it's upstream from one of the genes from the first gene that was on the screen.

Another zoom-out by 100-fold and then another 10-fold gives us all of the upstream regions focused in the center of the screen. We can now zoom back into that region and we can see that we have each of the items near the upstream region of the gene.

To get a closer look, let's condense the new track we just created using the right mouse button, to "dense" visibility and we will zoom into the 5-prime region of the two genes we discussed earlier, NBAS and DDX1. And here you can see there's an item upstream from each of the isoforms of each of the genes including the shorter version of the NBAS gene.

Now to intersect this with the SNP track, you want to zoom out and get the same region that we were looking at earlier so now this large 100x zoom shows us many, many genes and we will zoom in just to the region that has the upstream regions that we created in our new custom track. So once again we will go back to the Table Browser and we will confine our work to this particular region.

[5:23] Going back to the Table Browser now we will do an intersection and the Browser gives us the option of using either one of our two tracks that we are going to intersect as our primary table. Because we're interested in the names of SNPs we want to use the SNP track itself for our primary table so we will chose that as our table. We'll go to the "variation" group and then we will chose All SNPs from version dbSnp 138. Once again we will be sure we have the [position] button clicked and we will create an intersection using the "create intersection" button.

We want to go back to the custom track we created and use it as the secondary track. We have only one track in our custom track list and the Browser gives us a choice, "all SNP records that have any overlap with upstream200" and that's the choice we will accept. And then using the [submit] button we will go back to the Table Browser.

Here is where we will choose our output data format. Once again we will choose "browser extensible data" and we will [get output]. So let's include a custom track header here as well so that we can see on the Browser the results of our work. So we will call this "upstream SNPs" and we will "get one BED record per SNP" and then the "get BED" choice gives us our SNPs.

[6:43] Let's once again "select all" on this page and copy and we'll go back to the Genome Browser and load this page in as a custom track. So now you see that

because we have a custom track, our button has changed from [add custom tracks] to [manage custom tracks]. We will click this button and then we have one more step to go to [add custom tracks] to get us back to the familiar page where we paste our content. And once again we will zoom up to the top and we see that the information we input earlier is available to us as a track header. Submitting at this page takes us back to the [manage custom tracks] page and now we will click on the first of the upstream SNPs and it will take us to the Browser graphic. So here is the one base of the SNP and here is the 200-base upstream region that includes it.

So 100-fold zoom-out and another 100-fold zoom-out and then another 100-fold zoom-out takes us to a view where we can see that our SNPs align nicely with the custom tracks for the 200 bases upstream from genes. So let's pan over to the right a little now and we will pick up the DDX1 gene as well and then we will collapse the SNPs back down to "dense" visibility and you can see there are SNPs in each of the regions that we chose for our upstream regions.

S