How do I find all the SNPs in a gene using the UCSC Genome Browser?

This tutorial will demonstrate how to find all the single nucleotide polymorphisms in a gene using the UCSC Genome Browser.

To start we will go to the Genome Browser at genome.ucsc.edu and in the upper left-hand corner are two links that take us to the main page of the Browser where we can choose which organism to use. We will use the human default assembly hg19. The "`click here to reset`" button takes us to the default location and the [`submit`] button then will take us to the Browser graphical viewer.

[0:43] The default data are shown here, a number of data tracks. We will simplify the view by hiding all the data tracks and then we will turn the UCSC Genes track on to "`pack`" and hit the [`refresh`] button. The default gene here is the SOD1 gene. We will go to another gene, the SRC gene and choose that from the list and the [`go`] button will take us there. So the reason to show you this gene is to demonstrate that occasionally the canonical gene for a particular gene is not the largest transcript and if you're interested in getting the SNPs for all parts of the gene, you will want to widen the window just a little bit. We will nudge the window by opening it up using the little left arrow on the side here.

[1:35] Now the tracks for the single nucleotide polymorphisms are at the bottom of the screen and we will turn on three of those tracks: the Common SNPs from dbSnp version 138, the All SNPs track from the same version, and the Flagged SNPs track which are the clinically relevant or *potentially* clinically relevant polymorphisms. The refresh button then turns on all of those tracks to "`pack`", and we see that we get a very large number of SNPs with their rs numbers in the screen. The red ones are amino acid changes, the blue ones are splice sites or untranslated regions, the black ones are intronic regions.

Now the right mouse button will let us collapse the track down to a single row where it is "`dense`" and you can see that they're well distributed throughout the gene here. The All SNPs track is considerably denser that the Common SNPs track as you would expect and then finally the clinically relevant SNPs also include some green ones which are coding region SNPs that have no change in amino acid. We will collapse them as well down to "`dense`." Now they may or may not be clinically relevant and it is necessary to understand the exact nature of this track by reading the track description underneath the button on the left side. That's where you will learn that they are clinically relevant if they have a particular flag from NCBI where they're designated as having been contributed by a locus-specific database. This does not mean that they are necessarily disease-causing.

So now that we have seen the various types of SNP tracks there are, the All SNPs is the generally overarching track the other two are subsets of the all SNPs track. We will go to the Table Browser and use one of these tracks to find the SNPs in the gene.

[3:27] We will simply get the SNPs in this particular gene and the genome browser remembers where we were on the previous screen so that the coordinates are pre-loaded into this but it is necessary to click the [position] button. To find the SNP tracks we will select the variation group to match their location in the track controls beneath the Genome Browser graphical display. We will use the Common SNP track in this demonstration and we will get a list of SNPs. For "output format:" we choose "selected fields from primary and related tables" to get to a page where we can pick the columns we wish to save. And then the [get output] button takes us to a place where we can choose which features we wish to download as part of our file. On this page we can choose the name of the SNP. (We) can also choose the genomic coordinates if we're interested. For the moment we'll just choose the name. We can also choose a number of different items, such as the "function" and the "allele frequency" from this list. The [get output] button takes us to the list of SNPs.

[4:36] Now to demonstrate that we have the list we really want let's just choose just the rs numbers from the list, we'll leave the header behind, and we will copy this list back into the Table Browser and use it to generate a custom track so we can visualize our SNPs on the Genome Browser. And then the [Back] button takes us back and we will use the pull down menu to go to the main page of the Table Browser. Using the [paste list] button we will go to the page where we can load our SNPs into the Table Browser. The [submit] button loads the list of SNPs and then the "output format" button lets us choose "custom track" for our output.

The [get output] button takes us to a place where we can label our track. We will call it "common SNPS in SRC" and we will call the long label "common SNPs in SRC gene region" and use [get output] to see the custom track. And we will go over to the genome browser and view them. We can see the rs numbers and the tick marks corresponding to them. We will collapse them down into "dense."

Now the Common SNPs track can be moved up into place so that they are side by side with the new track we just created and we'll select this exon over here so that we can zoom into the region and get a closer look. You can see that there's a match between our newly formed track and the original Common SNPs track from which it derived. Now to see how we did restricting the range of our query, let's zoom out by a factor of 100 and we can see that all of the SNPs in our new track are situated over the SRC gene.