## How do I identify codon numbers with the UCSC Genome Browser?

This tutorial will demonstrate how to locate amino acid numbers for coding genes using the UCSC Genome Browser

First we will navigate to genome.ucsc.edu and arrive on the main page at which place we can read information about the Browser and recent news.

[0:36] We will use one of the links in the upper left-hand corner to navigate to the Genome Browser and this gateway page gives us access to a large number of other animals including the Human genome. We will click to reset and accept the default location at the human hg19 genome. In scrolling to the right side it reveals the "`submit`" button through which we gain access to the main Browser graphic. The Browser graphic has a large number of data tracks and defaults to the location at the SOD1 gene. To simplify the view we will simply turn all the data tracks off using the "`hide all`" button and begin by turning on a number of different gene tracks.

[1:35] The UCSC Gene set we will set to "`pack`", the RefSeq Gene set we will also set to "`pack`." This is produced at NCBI. We will also turn on the latest version of the GENCODE set which is produced at the EBI in the UK. The "`refresh`" button allows us to turn on each of these data tracks and you can see that each of the three annotation sets has similar notations for the exons and the introns in the gene. The SOD1 gene has a single isoform.

[2:17] So for the sake of looking at a more complicated gene let's navigate to the TP53 gene. Chose the gene from the list in the dropdown menu and then use the "`go`" button to navigate there. At the TP53 location we can see that there are multiple isoforms including a number of different start sites and in several locations, exons that are in one isoform and not another. The various gene sets have different numbers of isoforms and have various transcription or translation start sites. Let's zoom into these exons in the middle of the page by dragging our cursor in the position box at the very top of the page and then releasing it. This allows us to zoom into that location and at this position we can see that the amino acids are shown as alternating light and dark stripes but we are not yet zoomed far enough in where we can actually see the amino acid names.

[3:26] You also notice that the RefSeq genes track does not have the amino acids shown by default but we can use the configuration option to turn on the track coloring "`using genomic codons`" as the key and then the "`submit`" button returns us to the browser graphic and now the RefSeq genes track also has the alternating colors.

[3:51] Let's zoom once more into a closer view of the genes and at this location we can see the amino acid names have shown up in the single letter code and the amino acids have agreement in all of the isoforms which shows that each of these isoforms is using the same reading frame for all three gene sets. Now to view the codon numbers we can go back to the configuration page.

[4:23] Let's go to the "configuration" page for the UCSC genes track. We will click the show codon numbering checkbox and then the "submit" button to return to the Browser graphic. Now we can see that the same amino acid, for example this histidine, is numbered as 179 or 20 or 140 or 47 or 86 depending on the isoform we choose. So if you are interested in following the literature and using amino acid numbers that have been reported in the literature, it would be good to know which isoform was being used. Let's also do the same thing for RefSeq and we'll turn them on in the same fashion by "show codon numbering." Pursuing the configuration on the GENCODE track we'll also click into the minibutton.

[5:16] And we can see here that we're confronted with a different interface because this track is actually a composite track where only one of the five sub-tracks is turned on at the moment. And the track that is turned on is the "Basic" track which can be configured using the wrench sitting next to the pull down menu and the "show codon numbering" option is available on this page. We will leave that turned off for this track.

[5:44] So let's scroll down below the browser graphic to the track controls to the "Phenotype and Literature" section and turn on the OMIM Allelic Variants SNPs track to "pack" and we'll also turn on the UniProt Variants track to "pack" and use the "refresh" button to return to the Browser graphic with these two tracks turned on. You can see in this particular view OMIM Allelic Variants SNPs has one SNP. It's limited to a single nucleotide and it indicates an arginine 175 to histidine missense mutation and it also has an rs number, 28934578, [which] refers back to the dbSnp record. Now if we scroll down a little bit to see the rest of the changes that are reported in the UniProt track you can see that there are half a dozen different changes indicated in this track but the resolution of this track is limited to a specific amino acid rather that a specific nucleotide.

Let's move the screen over a little bit to the left so that our mouseover is visible for this change simply, by dragging. And now we can see that when we put the mouse over the UniProt record we can read that this particular change is associated with the Li-Fraumeni syndrome; germline mutation and in sporadic cancers.

[7:14] If we click into the record that matches between the two, that is to say, we click into the one that changes the amino acid to a histidine we will see that we also have dbSnp record in the UniProt track which matches the record in the OMIM Allelic Variants track. So let's go back now to the Genome Browser and confirm that at least one of the transcripts has 175 listed for this amino acid and you can see that we have one, two, three in the RefSeq genes track and also five of them in the UCSC Genes track.