

## How do I get the coordinates and sequences of exons using the UCSC Genome Browser?

In this tutorial, we will show how to obtain the coordinates and the sequences of exons using the UCSC Genome Browser.

To begin, we will navigate to the browser at [genome.ucsc.edu](http://genome.ucsc.edu). The main page of the Genome Browser is mostly a news page and information. In the upper left hand corner, you can find two links that lead to the Browser database.

[0:33] We will accept the defaults and use the human hg19 database. “`Click here to reset`” returns all the tracks back to the default locations and the default data. Using the “`Submit`” button, we access the main browser graphic. The main browser graphic page shows the default location, the SOD1 gene, and a number of data sets are turned on. You'll notice that the large boxes represent exons which will be our goal to obtain those coordinates and the thin lines are introns.

At the bottom of the browser, we can use the “`hide all`” button to eliminate all of the extra tracks and then turn on the UCSC genes track to “`pack`.” Using the “`refresh`” button, we turn the track back on.

[1:26] To get to a larger region that has multiple genes in it, we have a number of different ways to navigate. We can click on the coordinates and edit the coordinates in place. We can also type in information such as a cytoband (a chromosome gene band) such as 14q12 and use the “`go`” button to navigate to that location.

We can also use the navigation bars on the chromosome ideogram, going back now to the SOD1 gene, we can simply zoom out using the drag method around the red bar on the ideogram and the “`ok`” button gives us access to a small number of genes.

Let's zoom out a slight bit more, get a larger number of genes on the screen, and we will use these genes for our targets to get the exons.

[2:20] Using the pull-down menu at the top of the graphic, we go to the Table Browser. The Table Browser defaults to the genes and gene prediction group, the UCSC genes track, and the primary table of known gene for that track.

Note that it's also possible to use a different reference gene set such as RefSeq genes or the GENCODE gene set. We will use the “`UCSC Genes track`” and we will use the position information given to us that matches the information on the browser graphic because download from the whole genome will take a while and for the purposes of this demonstration, it's sufficient to simply use a confined region.

The Table Browser gives us a choice of a number of different output formats. We will start with the “`BED (Browser Extensible Display)`” output format which will give us the genome coordinates. The “`get output`” button gives us a page where we can choose a number of

different options. We will take the “Exons” option, we will include a custom track header, and we will call this "exons chromosome 21" and the long label we will call "exons chromosome 21 selected genes" and “get BED.”

[3:36] This page gives us the coordinates from the region we've selected, chromosome 21. It gives us our items, including each one with an exon numbering that matches the exons in the browser graphic, starting with exon-0. We also have the track header we chose back at the Table Browser interface. Selecting everything on the screen, copy. We can go back to the Genome Browser and view the data as a custom track.

[4:06] The “add custom tracks” button below the Browser graphic leads to the input page where we can simply paste the data that we obtained from the Table Browser and “submit.”

We now have one item in our "manage custom tracks" page and clicking on the first item in the "position" column returns us to the Browser graphic where we see the first item. We can see that this item is a single exon gene and to view the other items in the browser, let's zoom out by a factor of 100, and then 100 once again, where we see each one of our individual exons on the screen. To collapse the track to a more manageable size and turn off the names of all the items, we will select our visibility “dense” and we can see the custom track at the top of the screen with the genes in the view.

Let's drag our browser over to the left to center the custom track. You can see that we have the custom track in the screen limited to the positions where we were on the browser graphic in the first place. Let's select just that region and zoom in.

So this confirms that the coordinates that we obtained in the Table Browser match both the position and just the exons of the genes. You can see here that the individual items of the custom track line up nicely with the gene's exons, the TIAM1 gene in this case.

[5:40] Okay, we will now go back to the Table Browser to get the sequence of the exons. We will choose the default position which matches the location that was in the window of the browser graphic and let's change our output format to “sequence.” You'll notice that the Table Browser now uses the custom track as the default so we must go back to the “Genes and Gene Predictions” and we get the UCSC genes track and the knownGene table.

Okay, using the “sequence” output, the Table Browser can be used to get the actual sequence. We will use the “genomic output” option, and go to a page where we can unselect the introns. We can choose “One FASTA record per region,” and we can view the coding sequence in uppercase and the untranslated region in lowercase.

So here we have a page with our genes from chromosome 21. We have the output sequence with uppercase exons. You can see that the first item that was in our BED track earlier is a single exon gene. You can scroll down the page to see other genes, the coding regions of

which are broken up. Here is a region where the exon only because it is the third exon, exon\_2 of this particular gene.

[7:00] You'll also notice that the strand is negative. You can use this one as a test case to be sure that we're on the correct strand, that we have the actual sequence of the exon as it corresponds to the messenger RNA on the reverse strand rather than just the genomic coordinates.

We will copy it. We will go back to the browser, and using the "BLAT" function, we will paste this particular region into the BLAT and "submit."

So the first item in the list is a very good match, and we'll look at the "details" page to determine that we are on the correct strand. Going down to the bottom of the page using the "together," you see the coordinates of our pasted exon start with 1, and go up to 50.

And on the other strand, our coordinates are ...4456 and ...4500 running in the opposite direction. This confirms that our sequences for genes on the opposite strand (on the reverse strand), are the correct sequence in the sense of the messenger RNA, not in the genomic sense.