

How do I find a list of genes in a genomic region using the UCSC Genome Browser?

This tutorial will show how to use the UCSC genome browser to find a list of genes in a given genomic region.

To navigate to the Genome Browser go to genome.ucsc.edu.

This page shows the main page for the Genome Browser and one of the two links in the upper left-hand corner will take you directly to the gateway page where one can choose the organism. We will “click here to reset,” which puts us all on the default page so we can all start at the same location, which is the human hg19 genome assembly. By hitting the [submit] button, we go into the default page which shows the region around the SOD1 gene with a number of data tracks turned on.

[0:54] The [hide all] button below the browser graphic gives us an opportunity to turn off all the tracks and clean up the image. We will now turn on the UCSC Genes track to pack visibility which turns on all of the genes in this particular region, which is just the SOD1 gene at the moment. Now, to put a larger gene region onto the Genome Browser, one can type in gene coordinates using the same format as you can see here: “chromosome colon start-coordinate hyphen end-coordinate.”

[1:31] Or one can type in a number of other identifiers: “4p16.3” chromosome band. The [go] button will take us to that location. You can see that there are multiple genes in the window and multiple isoforms for many of these genes.

[1:54] To turn off the extra isoforms, go to the configuration page for the UCSC genes track by clicking the little button on the left-hand side. The “splice variants” box can be unchecked and then the [submit] button takes us back to the browser with only one isoform for each gene selected.

[2:20] To get a list of the genes in this particular region we will use the Table Browser by going to the “Tools...” menu then selecting the “Table Browser.” The UCSC genes track is the default location for the Genome Browser and we will accept the “knownGene” table. The “position” box can be selected here to limit our search to just the coordinates for the chromosome band identified on the previous page. The Browser remembers where you were in your previous browser window. For the “output format:” to get the gene names we must select from multiple tables so we will use “selected fields from primary and related tables” and then choose the [get output] button.

On this page we have the opportunity to choose “chromosome,” “transcription start” and “transcription end” and then from the kg cross-reference (kgXref) table we will choose “geneSymbol” which is the HGNC symbol, and a brief description of the gene. [get output] will give us a screen that shows us the gene names and the gene coordinates and their brief description. You can see that we have multiple genes

with the same name. So we're picking up each of the isoforms that were shown on the original screen.

[3:40] Going back to the table browser now, we can use a different table as the primary table if we are interested in getting only a single isoform for each gene. That table is the "knownCanonical" table. By typing the character "k" it takes me down into the k's and then the knownCanonical table can be selected from this list. Once again we will use "selected fields from primary and related tables." And the [get output] gives us a similar view where the knownCanonical table is the primary table. We will choose "chrom, chromStart, chromEnd," and then once again from the cross reference table "gene symbol" and "description." The [get output] button now gives us a page that has one isoform for each gene.