Making links to the UCSC Genome Browser. Part 2: Links inside genes and more tricks

Welcome to Part Two of the video series showing how to make links to the UCSC Genome Browser.

In the first installment of this series, we saw how to pick apart the UCSC Browser URL, how to turn on specific tracks of interest and control their visibility level, how to hide undesired tracks and how to navigate to locations in the genome using gene names or dbSNP identifiers.

In this, the second of the series, we will look at how to navigate to specific locations *within* genes without knowing their genomic coordinates, how to set highlights, how to find obscure configuration parameters, and how to use some other useful features for making customized links to the Browser.

In Part Three, we will discuss how to load composite tracks, how to access your remotely-hosted custom tracks and hubs and how to make useful links in your spreadsheets.

The distinctions between making direct URL links to the Browser vs Saved Sessions are discussed in a blog on the UCSC site, https://bit.ly/sharingBlog, in the first video of this series, https://bit.ly/ucscVid21, and in a previous video in the Genome Browser video collection: https://bit.ly/sessionVid.

[1:20] The basic URL

Let's start by building a URL using the principles we discussed in Part One and jump right into the Browser. We'll start with the basic URL, add a database/assembly name, hg38, turn off all default tracks with the hideTracks parameter and turn on the GENCODE Genes track to "pack" using the knownGene table.

https://genome.ucsc.edu/cgi-bin/hgTracks?db=hg38&hideTracks=1&knownGene=pack

Because this URL has no position information, it will go wherever you may have been in the genome before. I have set my Browser to the default location, but you may end up anywhere using this link. Not to worry. The next step will be to specify a location.

[2:01] Link to a coding nucleotide

It is possible to designate a specific nucleotide or amino acid of a protein in a URL. The Human Genome Variation Society has published standards for identifying deviations from the reference assembly. While there is only one "correct" HGVS designation for a variant, for your convenience, the Browser supports several different forms of modified HGVS nomenclature as well. A list of HGVS designations supported by the Browser can be found at https://bit.ly/ucscHGVS

Because most human genes have multiple isoforms, sometimes the designation of a specific amino acid or nucleotide by gene name and amino-acid number is ambiguous. The most reliable way to get the exact location is to use a RefSeq identifier, which is isoform-specific. RefSeqs are in the form "NM_<some digits>." For example,

NM_000257:c.1200C>T

is HGVS that identifies the 1200th nucleotide of a coding DNA, which will be in the 400th amino acid. Let's add that to the URL as a key-value pair specifying location:

&position=NM_000257:c.1200C>T

And we'll drop the hideTracks parameter.

http://genome.ucsc.edu/cgi-bin/hgTracks?db=hg38&knownGene=pack&position=NM_000257:c.1200C>T

The Browser will find the location and highlight the nucleotide, with five bases of padding on either side. Notice that the RefSeq track has turned on because the NM_ identifier identifies that track, so there are now two gene-prediction tracks in the image.

Notice that the amino acid numbering shows that the transcript is on the bottom strand, reading right to left, so that the highlighted *genomic* nucleotide is a G, not the C specified in the HGVS (which was C to T), which is transcript-oriented. You can reverse the DNA sequence in the image by using the little arrow at the left side of the Base Position track at the top of the image.

Let's click into the configuration button at the left side of track in the image and turn on the RefSeq accession numbers to confirm that we are truly at NM_000257. You see on the configuration page that there are other RefSeq tracks available in this composite track; but they are not turned on by default. Only the RefSeq Curated track is turned on. Returning to the Browser image, we see the expected NM number.

[4:53] Link to an amino acid

Let's try another variation of modified HGVS, in which we just use the gene name and designate an amino acid by its number. We'll go for amino acid 92 of the interleukin 6 gene, IL6, using

&position=IL6 %20 p.92

where %20 represents a space

http://genome.ucsc.edu/cgi-bin/hgTracks?db=hg38&knownGene=hide&position=IL6%20p.92

The "p." is actually optional, but recommended.

Notice that we are at leucine 92 in one of the isoforms, and the Browser is still showing the negative strand. We did not designate any change for strand in the URL, so it simply added our choice to the parameters already on display. Let's switch it back.

You see that the amino acid we have found is also called leucine 16 and leucine 69 in the other isoforms, where there are splicing events upstream that make the protein shorter. If you do not identify a specific RefSeq in your link, the software must choose one.

Each gene has an isoform designated a canonical transcript that is used in a case like this. The hg19 and hg38 assemblies use different methods to identify the canonical transcript for a gene. Details can be found on the configuration page by clicking the bar at the left side of the hg38 GENCODE track.

[6:33] Link to a coding nucleotide using gene name

We also use the canonical transcript when using "c." nomenclature to define the location in the coding DNA, such as

&position=IL6%20c.270

The "%20" is actually optional as well. You can use a space, as in the following URL :

http://genome.ucsc.edu/cgi-bin/hgTracks?db=hg38&knownGene=hide&position=IL6 c.270

This format does not specify a variant (as did the C>T that we saw in an earlier example).

And you can use a colon there in place of the space.

Using c.272 this time:

&position=IL6:c.272

http://genome.ucsc.edu/cgi-bin/hgTracks?db=hg38&knownGene=hide&position=IL6:c.272

Note the additional highlighting in the Browser image for each successive link.

HGVS also works on non-human assemblies. If you have a RefSeq identifier, it will work. E.g., this one for mouse:

http://genome.ucsc.edu/cgi-bin/hgTracks?db=mm10&knownGene=pack&position=NM_029083:c.510

goes to the mouse assembly mm10 in the Ddit4 gene at the designated nucleotide.

[7:44] Link with Short Match

Now let's look at the Short Match track, available in every genome assembly, which finds in the window a string of nucleotides from 2-30 bases long. We will modify a link that we used in the first video to go back to hg38, but at the helicase gene, DDX21, using:

&knownGene=pack
&singleSearch=knownCanonical
&position=DDX21

http://genome.ucsc.edu/cgi-bin/hgTracks?db=hg38&hideTracks=1&knownGene=pack&singleSearch=knownCanonical&position=DDX21

Then we'll add to the URL a parameter that turns on the Short Match track, finding all locations for the nucleotides GGATCC:

&oligoMatch=pack
&hgt.oligoMatch=<the sequence>

We'll discuss later how you might discover a feature like this on your own, but let's just tack it onto the end of the existing URL. You see that there are several occurrences of this oligonucleotide in the DDX21 gene region. This sequence is a palindrome, but short Match would find hits on both strands if were not. For example, if we drop the last nucleotide.

[9:12] Links with highlights

What if we wanted to highlight the 6th exon of this gene in a bright red color? First, we'll get the color. Use the mouse and select the region in the Base Position track at the top of the image. You can get the R,G,B hexadecimal color code right here in the color picker:

#ff0000

where

red = 255 or hexadecimal FF; green = 00; blue = 00

Choose the highlight option.

Remove the highlight using the right mouse button menu. Now let's get the coordinates using that same menu. Select the coordinates as before. And zoom. Click the coordinates on the left above the image and copy them from the Position Box:

chr10:68,966,695-68,967,475

The syntax for making highlights in links can be found on one of our help pages, where we describe the optional parameters than can be used in the header lines of custom tracks:
https://genome.ucsc.edu/goldenPath/help/customTrack.html#optParams

This page can be found in our help documents. Go to: Help > Browser Documentation. Then type "optional parameters" in either search box.

The syntax for a highlight is:

&highlight=
<db>.<chrom>:<chromStart>-<chromEnd>#color

The colon and hyphen may be encoded in a URL or not. The hashtab *must* be encoded or it is ignored, because a bare hashtag will be interpreted by your web

browser as an instruction to jump to an anchor tag in the page, the way the "#optParams" anchor jumps to the middle of the help page seen earlier.

&highlight=<db>.<chrom>%3A<chromStart>%2D<chromEnd>%23<color>

colon is %3A
hyphen is %2D
hashtag is %23

So let's encode our highlight on exon 6 to be green:

We'll add

%2300FF00

and use a slightly rounded set of coordinates:

&highlight=hg38. chr10:68,966,700-68,967,500%2300FF00

Removing the "%23" and putting the hashtag in the link causes the color parameter to be ignored and we get the default highlight color.

&highlight=hg38. chr10:68,966,700-68,967,500#00FF00

Let's remove the highlight and test this. This demonstrates that the hashtag must be encoded to get the color you want.

Making the full URL as shown:

http://genome.ucsc.edu/cgi-bin/hgTracks?db=hg38&hideTracks=1&knownGene=pack&singleSearch=knownCanonical&position=DDX21&highlight=hg38. chr10:68,966,700-68,967,500%2300FF00

will show the whole gene and highlight exon 6 in green.

We have hideTracks, we have knownGene=pack, we have singleSearch=knownCanonical, and position=the gene DDX21, and then the highlight.

This gives the full gene in the window and the exon highlighted green.

While this video was in post-production a change was made in the format for highlighting. When the change is released in a few weeks, all of the separators between data items will be hashtags (%23 in links).

There are a number of other options that can be incorporated into links that have not been presented here. For example, if you do not like the padding options we have seen, such as five bases on either side of an HGVS nucleotide or amino acid and 250 bases on either side of a dbSNP accession (which was demonstrated in the first video of this series), you can encode the action of the buttons above the Browser graphic and zoom in or out as indicated. These options are detailed, along with several other options, on the optParams help page mentioned earlier:

https://genome.ucsc.edu/goldenPath/help/customTrack.html#optParams

There you will also see some of the features we have discussed in these videos, including highlighting.

[13:27] Finding other options: The cart

There are situations where you may have a configuration on your screen, displayed through tweaking various buttons and pulldowns in the interface. If you wish to capture the configuration and use it in a URL, you can learn the syntax for the parameters by dumping a list of the settings to the screen using the cartDump cgi:

https://genome.ucsc.edu/cgi-bin/cartDump

This will show all active variables and their values.

Let's dump the current cart and see what we have. Several of the settings represent familiar settings from the first video in this series, such as the current genome,

db hg38

default tracks that have been turned off: 100-species Conservation track, hide;

cytoBandIdeo hide
gtexGene hide

and a highlight.

#00FF00

You recognize green there.

The current width of the screen is recorded:

pix 1027

And it remembers the Short Match setting from earlier, even though the track is turned off:

hgt.oligoMatch ggatc

You can use any one of these key-value pairs in a URL.

You can experiment with the variables right at the bottom of the cartDump page to see what they do. An obvious one is pix. Let's set it to 1600, which will be pretty obvious back on the graphic and [submit].

To get back to the graphic, hack the URL by adding "hgTracks" after the cgi-bin in the URL. You may remember that as part of the base URL described in the first video of this series. All of the parameters seen in cartDump will be used to configure the graphic and you see that the image is indeed really wide.

We can reset the window using the resize button below the Brower graphic, to match the window size, or we can set it with the URL using the original value in the cart:

&pix=1027

appended to the URL.

It is worth mentioning that the cart can also be exported via the Saved Sessions page. Go to My data > My Sessions and, in the "Save Settings" section, choose the option to save to a file.

[16:01] View Browser graphic only: hgRenderTracks

One interesting variation on the URL is the situation in which you may wish to display or embed just the graphic image from the Browser without all of the track buttons, documentation and track controls. In that case, use the cgi:

hg*Render*Tracks

instead of

hgTracks.

For example, let's modify the URL we just used by adding "Render" to the URL.

We'll add the new encode cis-regulatory element track, just for fun, using:

&encodeCcreCombined=pack

http://genome.ucsc.edu/cgi-bin/hgTracks?db=hg38&hideTracks=1&knownGene=pack&singleSearch=knownCanonical&position=DDX21&encodeCcreCombined=pack

This gives you the Browser graphic only. Because the image is no longer embedded in a page with the coordinates visible, they are added to the top of the graphic. Note that highlights cannot be displayed as part of hgRenderTracks.

That concludes Part Two of the three-part series of videos showing how to make links to the UCSC Genome Browser. You can find more information about making URLs to the Browser in a blog written by Christopher Lee of the Browser staff at:

https://bit.ly/urlBlog

In Part Three of the series we will discuss how to load composite tracks, how to access your remotely-hosted custom tracks and how to make useful links in your spreadsheets.

You can subscribe to the UCSC Genome Browser YouTube channel and watch for new installments at

https://bit.ly/ucscVideos

This video was released during the global coronavirus pandemic. Please note that when it is safe to travel again, we will restart our on-location training program. You can contact us to host a one-day or two-day in-person training at your location at

https://bit.ly/ucscTraining

We are also open to Webinar-style instruction.

Thanks for watching and thanks for using the UCSC Genome Browser.