UCSC Genome Browser Basics. Part One: Getting around in the Browser

Welcome to Part One of the Basic Browser video series. We will focus primarily on the use of accession names – the many ways that identifiers of various kinds can be used to locate genomic locations and annotations. We will also show a few configuration options for setting up the Browser. In the next installment, we will focus on more ways that the Browser image can be configured to show the data the way you want. In Part Three, we will show you many more configuration options and also how you can use DNA sequences to navigate. In this part, we will deal mostly with navigation and how to get around in the Browser.

[0:40 Set Browser to hg19]

To start with, let's set our Genome Browser to the default settings, "reset all user settings," and then from the gateway page, we will choose the human hg19 genome. That genome assembly is still the most popular genome among the human genome assemblies. Hit the "go" button to get into the Genome Browser.

We have a large number of data tracks turned on by default in the browser graphic. We define "track" as set of data with at least one database table underlying it. Typically all data in such a table will be of a similar type, such as gene predictions, mRNA mappings, or transcription-factor binding sites. Let's simply turn them all off; we'll hide all of the browser tracks using the "hide all" button.

[1:32 Navigate by gene name]

To find a place to start with let's simply type in a gene name in the Position Box. Let's go with EGFR, and we notice that it comes down in the drop-down menu as one of the options. We can select that and hit "go" and the Browser will navigate to the EGFR gene and turn on the appropriate gene track. In this case, the UCSC Genes track.

[2:00 Use gene name and amino-acid number]

There are other ways to navigate within in the browser to get to particular locations. For example, in the Position Box, you can type in the name of a gene and the number of a particular codon if you wish to navigate to that codon. For example, FGFR2 followed by a space and a "p" with or without a dot and "33" will take you to the 33rd amino acid of the FGFR2 gene. You'll notice that a second gene set has come on in this case, the NCBI RefSeq Genes. The nomenclature we use for the "p dot" syntax is keyed to the NCBI RefSeq gene set

and so whenever you use that type of navigation it turns on the RefSeq gene set. It is presented on the screen in the center with 5 bases on either side.

And you can see that sometimes the numbering is different: If there's been an upstream splicing you may get a different number for some of the isoforms. For example, leucine 33 here at the center of the screen is also annotated as leucine 52. You may notice that this particular gene is transcribed and translated on the opposite strand so the numbering goes from the right to the left. You can switch the orientation of the DNA at the top of the screen by clicking the little arrow at the upper left corner which reverses it so that the TTA codon that you see in the window reading from right to left matches what you would see in the codon tables.

I'll switch it back now to have it in the normal genomic order from left to right.

[3:40 dbSnp: Reference SNP (rsIDs)]

A number of other things can be typed into the Position Box, such as the reference SNP identifier of a particular nucleotide variant out of dbSNP. For example rs10000. If you type that and hit "go," you will get a number of different hits on this page, indicating that the particular variant is found in a number of different releases of dbSNP. The higher the number, the more recent the version of dbSNP. So, we'll just turn it on in the "All Short Genetic Variants from dbSNP 153" track and you can see we are going to go to chromosome 7. The SNP track comes up with the variant you have chosen highlighted in faint yellow coloring in the background. The label of the individual variant is written in reverse video where it's white against the colored background. In this case, it's a green background because the variant is a synonymous variant; it does not change the amino acid.

[4:47 RefSeq NM_ identifiers]

So let's continue to look at ways to get around in the Browser using various accession numbers, names, and identifiers, and to do that let's turn off, using the right mouse button, the "All SNPs" track and we'll also turn off the NCBI RefSeq track and leave just the UCSC Genes track turned on.

Another type of identifier that you can use is a RefSeq identifier. For example, NM_014877.4 and you'll notice once again the RefSeq track comes back on and we're navigated to the HELZ gene and it is highlighted. If there are multiple isoforms, the various isoforms will have different NM numbers and the one that you typed in is the one that is highlighted.

[5:37 OMIM identifiers]

The Browser also supports OMIM identifiers (Online Inheritance in Man), or example, 115500 and hit "go." It's the only type of identifier the Browser recognizes that has just integers and it looks it up in the OMIM table and you can see that it is highlighted here. The OMIM gene track comes on and we have navigated to the catalase gene.

Typically when any browser track comes on, and you click into the browser item, the item in the track takes you to a details page. In this case, you can click through to the gene at OMIM website, or to the phenotype record. You can see here the 115500 matches the identifier we put in, but there is also a phenotype identifier 614097 and if you click into that, it takes you to the OMIM page for Acatalasemia. We'll just close that window and navigate back to the Genome Browser.

[6:39 Genomic coordinates]

There are a number of other ways to navigate in the browser. For example, you can use genomic coordinates if you'd like. Click on these coordinates on the left to get them into the edit box. You can simply tweak the end coordinates by adding 100 on one side and subtract 100 on the left side and you would go from a 33,127 total window size to 33,327 and you've put a little bit of white space on either end of the gene.

You can also add a little white space on the end of the gene by nudging the edges here. If you click to the left on the "move start" it moves over a little bit. If you click on the right you can do the same thing and move it in a little bit. And if you want, you can change the number next to the arrows and it'll change the magnitude of the nudge. The number represents the number of light blue lines that you will move.

[7:37 Tweaking the coords]

So let's look at the coordinates again. We'll click over here in the box and put the coordinates into the edit box and go in and just remove the commas and hyphens and the colon so that you get the coordinates in a format that you might have seen in another application or you copy/pasted it from another application.

And in fact, it is now possible to remove the "chr" at the beginning of the chromosome name and simply enter the coordinates here. And so if you click "go" with those coordinates in the window, you see it goes to the same location as it was previously.

[8:18 Single-nucleotide coordinate]

It's also possible to put in just a single coordinate to go to single nucleotide. For example, you can type in chr11:34477700 and hit "go." That'll go to that single nucleotide. And that's the only nucleotide in the screen. What does not work is putting in just the bare chromosome number without the "chr" in this format. When you type in a single coordinate, the "chr" and the colon are required, but the commas in the number are not.

[9:02 Exon and codon numbers]

You'll notice that when you're zoomed into an exon far enough in, there's room enough on the screen to show you the amino acid and the amino acid number, then by default, the number gets turned on in a number of our gene tracks. You may also have noticed when we put the mouse over the exon that you get an exon number here for each one of the isoforms. You might get a different exon number for different isoforms at the same location. And it tells you how many exons there are in the gene and which exon your mouse is at. There's also a double-headed arrow at the end that takes you to the end of the exon or the start of the exon on either end of the screen. If you click that, it'll go all the way to the end of the current exon.

So here we are at the left-most nucleotide of this exon. You'll notice that if we zoom out by a factor of three to get a little more on the screen, we pick up the next nucleotide in each direction, and you can see that the intron right next to the exon is shown. The intron is shown as a thin line with arrowheads showing the direction of transcription and translation.

Let's go back to the position we were at before, on chromosome 11, 34477700 and then zoom out a few times. If we zoom out by a factor of 10, we're at 10-base window. And then another factor of 10: We're at a 100-base window. You'll notice that methionines in the browser are always colored green. Some of them are a start codon, some of them are not, but they're always green.

At this resolution there's no room to write the amino acid numbers, but the amino acid names are still there in single-letter code and then zooming out again by another factor of 10, there is now no longer room to put the amino acid symbols in there, either. And so it's just an alternating dark and light color indicating the codons. That saves you the trouble of having to subtract two numbers from the scale bar and then divide by three to estimate how many amino acids you might have in your window.

[11:13 Cytobands]

Yet another way to navigate in the browser is to use the cytological band nomenclature. For example, up here you can see we're on chromosome 11 and you can see that here's band p15.1. Before I use that, 11p15.1, I'm going to turn

off at least one of the gene tracks because we're going to get a lot of genes in the window. And then I'll hit "go," and you'll see that we are highlighting in red in the chromosome ideogram, the region of that cytological band. Cytoband nomenclature is useful when navigating on the human, mouse and Drosophila genomes, all of which have well-established cytology.

You'll notice that the chromosome band track came on, and so 11p15.1 is shown here. If I put my mouse in the window and drag it over to the left, then the next band over will here also show up in that track and you can see 11p14.3. So that's the next band closer to the centromere. It's worth mentioning that the p-arm is always on the left.

[12:28 Display single, canonical isoform only]

You'll notice when you zoom out that you get a lot of genes with a lot of isoforms in them and it tends to clutter up the screen. So it's useful to be able to turn those off if you are looking at this large-scale resolution.

For any one of our data tracks over here on the left side is a little gray bar that turns to blue when you put the mouse over it. Note that the background of the track is light green. Each track has a label in the middle of the image, even if there is no data in the window being displayed. If you wait long enough, the mouseover on this bar tells you that you can click it to get to the configuration page, and one of the configuration options for the UCSC Genes track is you can turn off the splice variants (on hg38, it's on the GENCODE Genes track). When you do that and hit "submit," it cleans up your screen a little bit, showing only the canonical transcript for each gene.

The button on the left side of the track that we used to access the track configuration option is not the only way to get to that page. If a track is not turned on, then it is convenient to use the configuration option available by clicking the link above the pulldown menu for any one of the tracks. If we click once again into the UCSC Genes track, you can see that it's the same page and it has the splice variants option.

And at the bottom of the page on any one of our tracks, you can see a Description that tells you why the data are useful, how they were developed and a bunch of other information such as color options and a list of references at the bottom of the page.

If I go now back to the Genome Browser, I can show you yet a third way to access those configuration options. For any items in the track, if you use the right mouse button you have the option to configure the track in a context-sensitive menu, except that in this case you do not get the track Description, but you do

get all the configuration options, the splice variants and so forth, that we saw before.

That concludes Part One of the Basic Browser Video series, which has primarily focused on the many ways accession names from a variety of datasets and data types can be used to find your way around the genome. There are other accession names that work as well. The best advice is to just give it a try in the Position Box.

We will start Part Two at the same genomic location. To make it easy to find that location again, we will save the configuration for easy access. The Saved Sessions feature is the subject of a separate video in the UCSC Browser video collection, so it will not be repeated here.

Please note that we conduct full-day and two-day on-site trainings at your institutions. Prices are reasonable. Visit

  http://genome.ucsc.edu/training/

Thanks for watching and thanks for being a UCSC Genome Browser user.