How do I learn which tables belong to a data track on the UCSC Genome Browser?

This tutorial will demonstrate how to learn which tables in the UCSC database are associated with the data tracks in the Genome Browser graphical viewer.

We will navigate to genome.ucsc.edu and on the main page of the Genome Browser there are two links in the upper left-hand corner which will take us to the main gateway page where we can choose our organism of interest.

[0:42] So we will work with the human genome and we will use the "`click here to reset`" button which takes us to the default genome assembly with a particular set of data tracks turned on. The "`submit`" button then will take us to the main graphical viewer. There are a large number of data tracks turned on by default on the Genome Browser and the most convenient way to deal with them is to hide them and then we'll turn back on the tracks we're most interested in.

[1:18] Let's turn on the UCSC Genes track to "`pack`" and scrolling down to the bottom of the track controls we can turn on a couple more tracks in the variation group, specifically the Common SNPs track from dbSnp version 138, currently the most recent version, and Database of Genomic Variants (DGV) Structural Variants. We'll turn them both to pack. You'll notice while we are here you can put your mouse over the title and get the longer version of the title. You can click any one of the "`refresh`" buttons then and the three tracks that we have selected will turn on and you will see that we are on the default gene, SOD1.

[2:03] For the sake of variety, let's move to a different gene, TP53, and we will choose it from the top item in the menu here and then use the "`go`" button to take us to that gene. The TP53 gene has a number of isoforms and you can see that there are also a number of SNPs in this region. The red ones indicate single nucleotide polymorphisms that change an amino acid and you can see that the Database of Genomic Variants track has two sub-tracks which contain items at this location.

[2:45] Now there are a number of different ways to determine which tables underlie these different data tracks. The first and most straightforward is to simply put the mouse over one of the items in the data track and you can see at the bottom of the screen that for the UCSC Genes track it says in the middle of the screen "`hgg_type=knownGene`." This is a little cryptic but it identifies for this track the identity of the table. Now most of the tracks on the Genome Browser follow the format "`g=`" and you can see in the middle of the screen at the bottom, "`g=snp138Common`." That's the name of the table which incorporates in it the name of the version of the database. The DGV track has dgvMerged and the other sub-track has a table dgvSupporting. Each one of these two tracks is underlain by an individual table.

[3:38] Now another way to tell the name of the table is to click on the little button on the left side of the screen. When a track is turned on, this little button takes you to a

configuration page which in the case of the SNP track here, takes us to a page with a large number of filtering options and a number of other choices. On this page the "view table schema" link will take you to a page that shows you the name of the database and the primary table that contains the coordinates for each of the items in the data track, if we wanted to download the table or query it using the Table Browser. Going back to the Genome Browser using the navigation bar at the top of the page,

[4:33] … we scroll down the screen to the track controls. We find that if we click into the link above the track control it takes us to the same page as does the minibutton next to the data track. This has the advantage that you can click into one of these track controls without having to actually turn the track on.

[4:47] So yet another way to find the identity of the table is to click into one of the items in the track itself. So let's click on this red-colored SNP which is an amino-acid-change SNP and you will find that in this location we get individual information about that particular SNP. If you scroll down the page you can see that there is a link on this page as well to the table schema.

[5:25] Click through to the table browser from this location and you will find that the Genome Browser takes the information for the track that you are viewing, if you are viewing detailed information, and loads it into the Table Browser choosing the correct track group "variation" and the correct track name and the correct table. Now you'll notice that there are a number of other tables that are associated with this track and these are all secondary tables that contain information that's linked back to the primary table. The primary table will have coordinate information and the secondary table will have information that is linked back to the primary table using one of the identifiers.

At the Table Browser we can download some sample data records from the table to see if it contains the information we are actually interested in. And it's best to actually download records from a small location, in this case the location that corresponds to the page we were looking at, at the Genome Browser graphical viewer. If you use the "whole genome" option which comes up by default, you will download a very large number of records and it may take quite a while. The "output format," "all fields from selected table" gives you a look at all the data fields in the record and the "get output" button then takes you to a page that gives you all of the fields for the table you are looking at.

[6:53] If you are interested in a subset of the fields, go back to the table browser and choose "selected fields from primary and related tables" and from there you will have the option to choose just which data fields you are interested in: "chrom, chromStart, chromEnd," and "name" for example, and then beneath the checkboxes you can "get output" and retrieve just the data you're interested in.

[7:25] If you *are* interested in downloading the entire table and the entire contents of the table you should go to the downloads section. Let's go back one, using the "back" button and we'll go back up to the top of the screen and use the "home" button which takes us

to the main page again and on this page we can choose downloads and the “downloads” button will bring us to this page. The Human genome is found at the top of the list and we can scroll down the page to the hg19 data set where we can choose the annotation database, and remembering the name of our file is snp138Common, we can simply scroll down the page using “snp138c” and you can download the entire contents of this table.