Tutorial

An Introduction to Annotation Tracks

March 31, 2016

An Introduction to Annotation Tracks

In this tutorial we will download an annotation file from an external source and import this as a track. Doing this allows you to add valuable information to an existing reference track in the *CLC Genomics Workbench*.

There is a separate tutorial called *Reference Genome Tracks* where in addition to describing how to get a reference genome track, we look at the **Download Reference Genome Data** functionality of the Workbench, which allows you to download both sequence and annotations for selected model genomes.

As a prerequisite to this tutorial, please download human chromosome 17 and convert this into a track. Instructions on how to do this is found in the tutorial *Reference Genome Tracks*, see *Method two: Using other sequence data sources*.

Downloading annotation data from external source

We will download *dbSNP(common)* annotations from the UCSC site. We are working with chromosome 17 of the human genome, so we wish to download annotations specific for this chromosome. If we download the dbSNP Common annotations via the **Download** function in the Workbench, we will get all chromosomes and to be able to download annotations for only chromosome 17 we have to do it directly from the UCSC site.

We need to download the annotations in a format that is recognized by the *CLC Genomics Workbench*. Here, we will choose txt format. To save space, we will choose to download it as a gzip compressed file.

The full list of annotation formats recognized by the Workbench can be found in the *CLC Genomics Workbench* usermanual.

1. Go to the UCSC table browser: `http://genome.ucsc.edu/cgi-bin/hgTables`.

2. Specify the search parameters as shown in figure 1. Key to downloading annotations only for chromosome 17 are:

   - Selecting **region: position**
   - Entering **chr17**
   - Clicking on the button labelled **lookup**

   Note: To download dbSNP for the entire genome, you should simply select **region: genome** instead.

   Remeber to give the output file the following name: "dbSNPCommon.txt", and to chose to download it as a gzip compressed. Click on "Get output".

   Once all the parameters are as shown in figure 1, then

3. Click **get output**.

4. Choose to save the file if you are prompted for this.

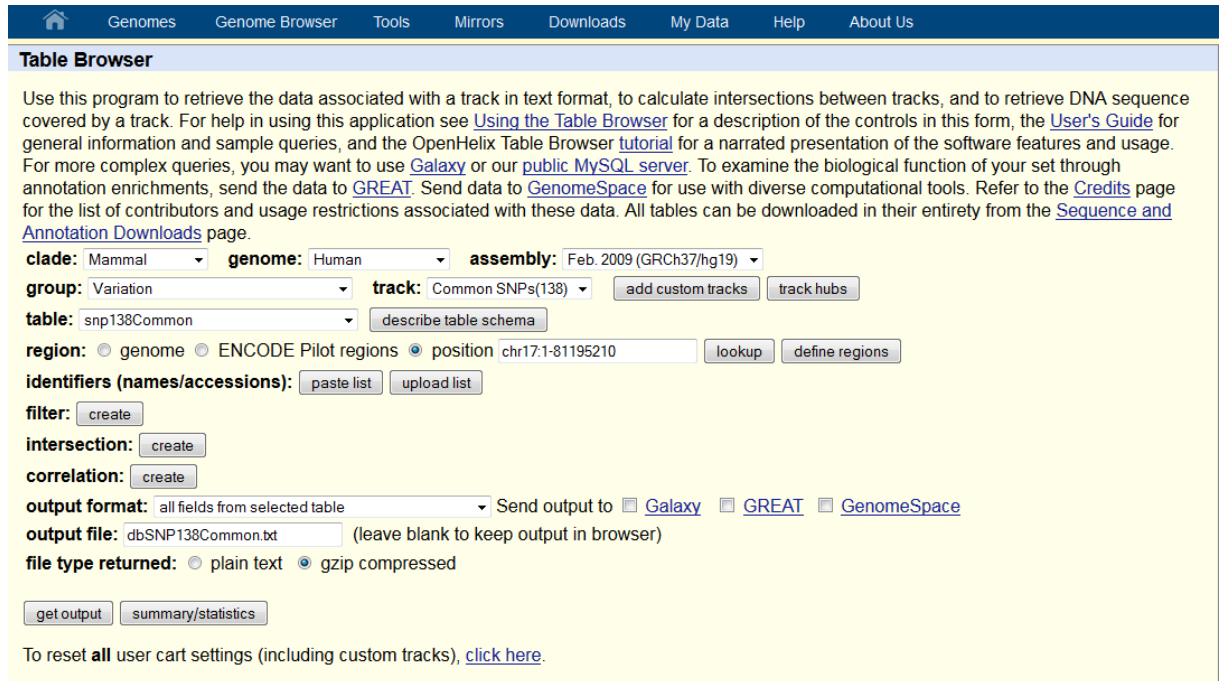If you have used the settings shown in figure 1, then the file you download will be called commondbSNP138chr17.txt.gz.

Figure 1: *Search terms for the UCSC table browser window.*

Importing the annotation data

Now we have a compressed txt format file containing the dbSNP common variations. We now import this data into the Workbench.

We assume here that you already have human chromosome 17 imported and have converted it to track format. Figure 2 shows what this data might look like in the Navigation area of the Workbench. Depending on how you obtained chromosome 17, the name of the sequence object may differ from that shown. For example, chr17 instead of NC_000017.

If you do not yet have chromosome 17 imported in track format, please refer to the tutorial *Establishing a Reference Genome Track*.
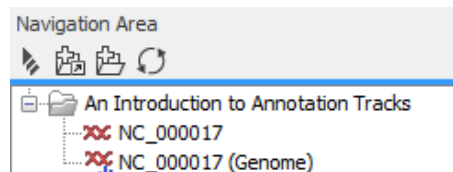

Figure 2: *The Workbench navigation area should look like this when you have downloaded the chromosome 17 sequence and have created a track version of the chromosome 17 sequence.*

To import the dbSNP common variation file:

1. Open your *CLC Genomics Workbench*.

2. Import the dbSNP annotations as a track by going to:

    Import ( ) |Tracks ( )

3. Under "Type of files to import" click on the drop-down list and select **UCSC Variation database table dump** (see figure 3).
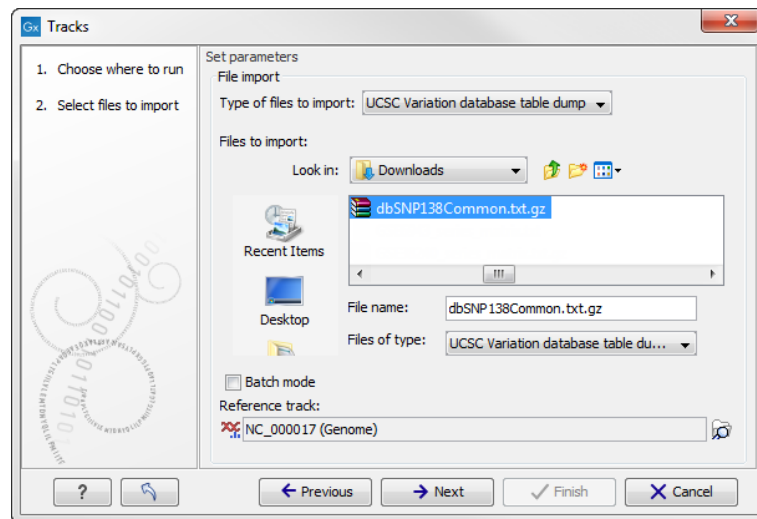
Figure 3: *When you import the dbSNP common annotations, the "Type of files to import" should be switched to "UCSC Variation database table dump". Find and select the annotation file that you have downloaded and under "Reference track:" select the chromosome 17 track. If you cannot se the chromosome 17 track in the folder in the Navigation Area that holds the data, it may be because you have not yet converted the chromosome 17 sequence to a track.*

4. Select the annotation file, which will be named **commondbSNP_chr17.txt.gz** if you used the settings from figure 1.

5. For **Reference track ( )** browse to locate the chromosome 17 reference track. For the data in figure 2, this would be the track called **NC_000017 (Genome)**.

6. Click on the button labeled **Next**.

7. Choose to **Save** the track.

8. Click on the button labeled **Next**.

9. Choose a location to save the track to.

10. Click on the button labeled **Finish**.

Importing can take some time, depending on the size of your annotation file.

As the name of the imported data is "Variants", it may be a good idea to rename to a more appropriate name.

Viewing the annotation track

Once the import has completed, you can open the track in the viewing area of the Workbench and investigate the annotations themselves.

1. Open the annotation track so it is visible in the viewing area of the Workbench. By default the track view is zoomed all the way out. That is, you see all the annotations for the whole chromosome. The annotations are represented by vertical bars.
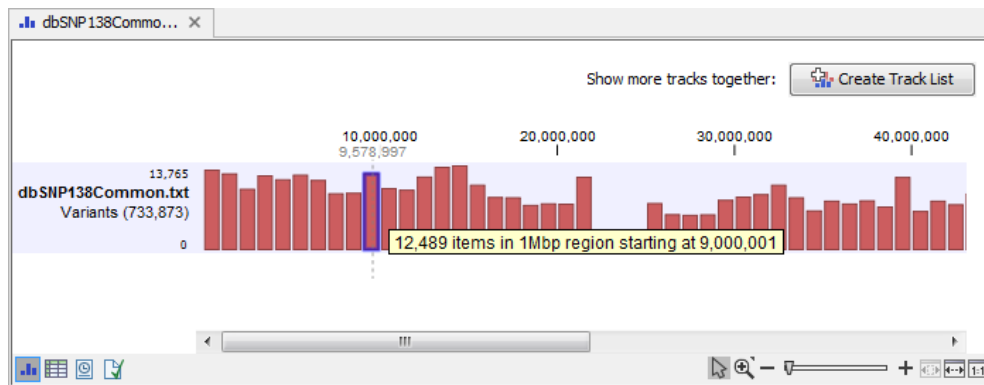
Figure 4: *Open the commondbSNP_chr17 track to have a look.*

2. Mouse over one of the bars. A pop-up window will provide information on the number of SNPs (items) in the specified range of bases (see figure 4).

3. Zoom in until each SNP is represented by its own small bar.

4. Mouse over one such bar. Now the pop-up window will display the information available for that specific SNP (see figure 5).
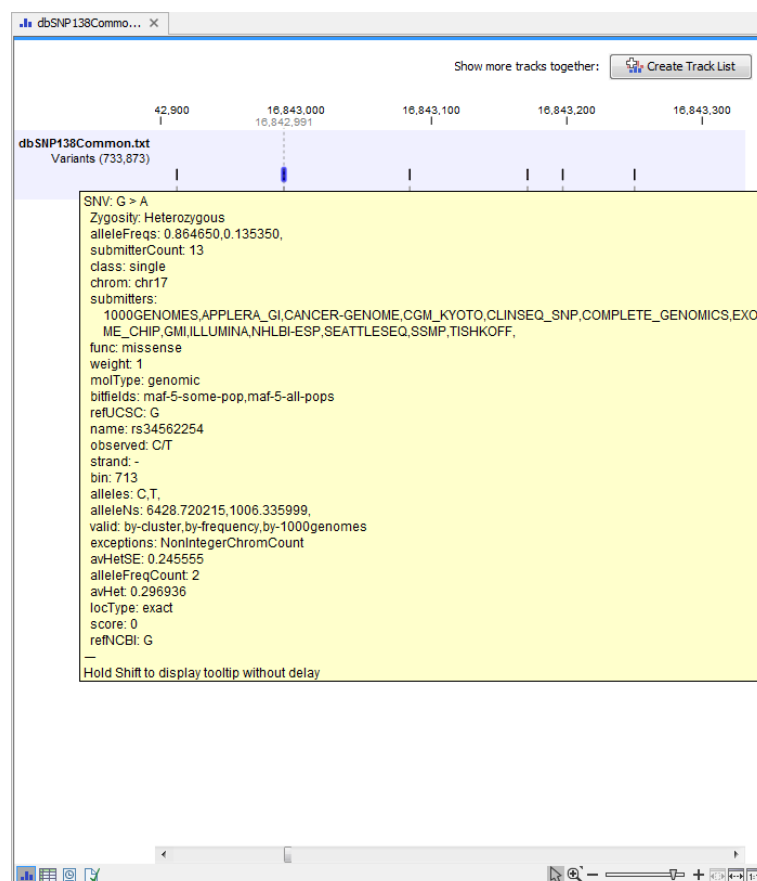

Figure 5: *Zooming in, individual SNPs are represented as bars. Mousing over results in a pop-up box with information about the SNP.*

You can also view the annotation track in table view. To open this in a split view:

5. Press and hold the Ctrl-button on the keyboard (⌘ on Mac) and click Show Table (▦) at the bottom of the view.

This brings up a split view showing the track and table view at the same time. Notice that the two views are linked. Selecting a row in the table highlights the specific variant in the track view with a vertical line (see figure 6).
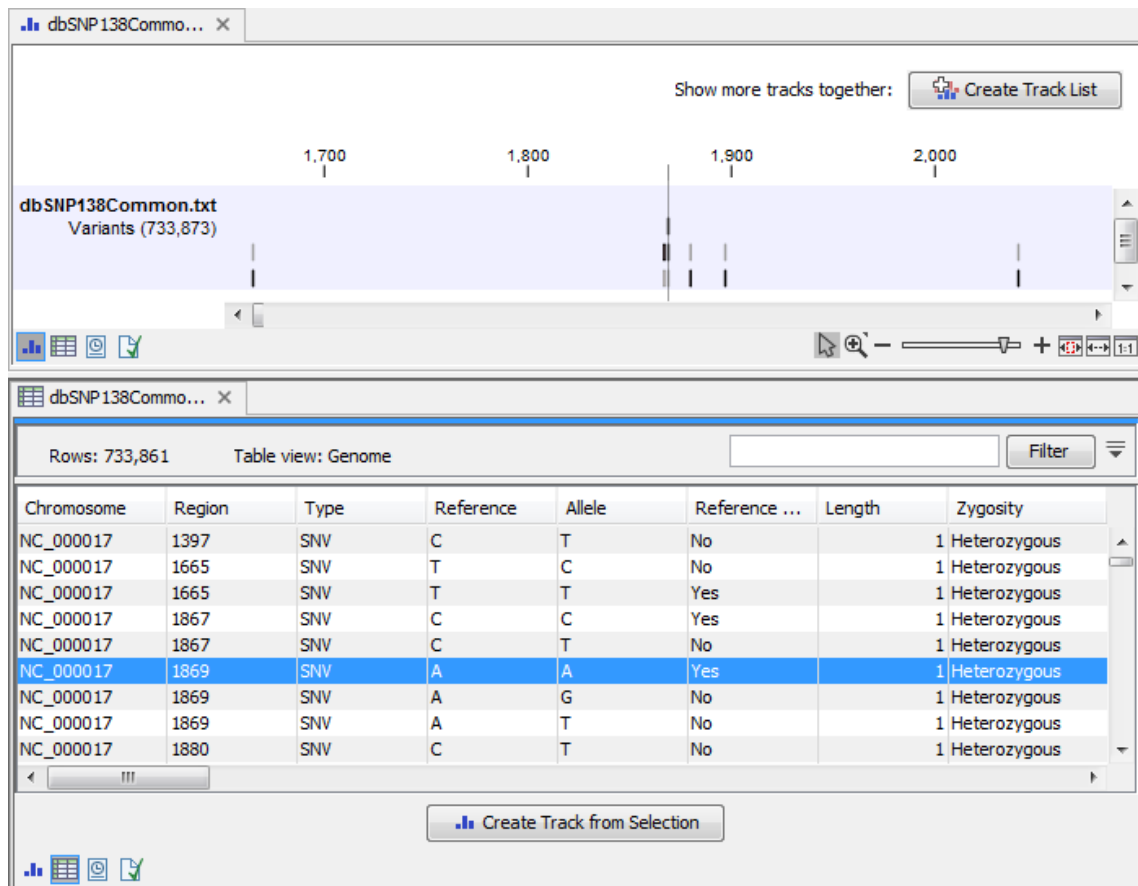
Figure 6: *Track and table split view of the annotations. The two views are linked.*

The table can be filtered in different ways. The most simple way is to click on the heading of a row. The other option is to use the filter found in the upper right corner of the table area.