



Tutorial

Find Actionable Variants With QIAseq Panels

June 6, 2017

— Sample to Insight —

Find Actionable Variants With QIAseq Panels

This tutorial uses the capacities of the Biomedical Genomics Workbench and the QIAseq DNA V3 Panel Analysis Plugin to find actionable variants - even at very low frequencies - in Targeted Amplicons sequences generated using a BRCA1 and BRCA2 QIAseq panel kit.

This tutorial covers in just a few steps all the following:


- Import Illumina paired reads in the workbench.
- Create a sequencing adaptor list to remove the sequencing primers off the reads with the Prepare Raw Data workflow.
- Download the hg19 Reference Data Set.
- Import the file listing the V3 primers specific to the panel used.
- Import the target region file specific to the panel used.
- Import a PCR universal adaptor list.
- Find low frequency variants with the **QIAseq Panel Analysis** Ready-to-Use Workflow.

Prerequisites

For this tutorial, you must be working with the Biomedical Genomics Workbench 3.5 or higher. You must also have installed the QIAseq DNA V3 Panel Analysis Plugin. How to install plugins is described here: http://resources.qiagenbioinformatics.com/manuals/biomedicalgenomicsworkbench/current/index.php?manual=Installing_plugins.html.

Download and import data

Go through the following steps to download and import the data into the Workbench.

1. Download the sample data from our website: http://resources.qiagenbioinformatics.com/testdata/QIAseq_V3_tutorial.zip and unzip it in the location of your choice on your computer.
2. Start the *Biomedical Genomics Workbench*.
3. Import the reads via the toolbar: **File | Import**  **| Illumina**
 - Select the two fastq paired files (figure 1).
 - Under "General Options" section, ensure the **Paired reads** and **Discard read names** checkboxes are checked.
 - In the "Paired read orientation" section, ensure the **Paired-end (forward-reverse)** option is checked.
 - Set the **Minimum distance** to 1 and the **Maximum distance** to 1000 (default values).
 - Click on the button labeled **Next**.

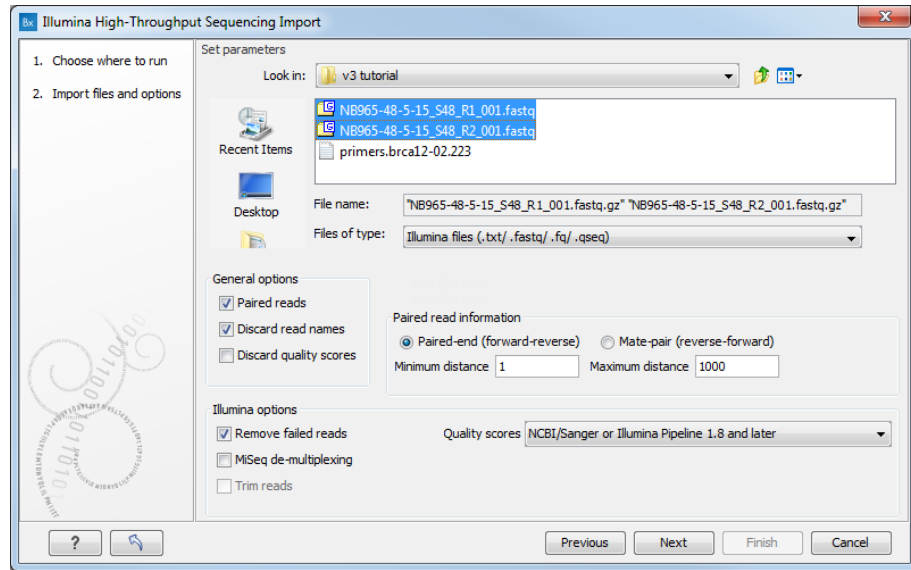


Figure 1: Import the reads.

4. Click on the button labeled **Save** in the wizard page that appears, choose the folder you wish to save the reads to (you can create a new folder dedicated to this tutorial for example) and click **Finish**.

Both files have been merged as a single file of paired reads during import.

Data management configuration

In order to do this tutorial, you need to use the hg19 Reference Data Set.

1. To download this dataset, go to:

Toolbar | Data Management (📁)

2. This will open the wizard shown in figure 2.

Select the Reference Data Set hg19 and click on the button labeled **Download**.

3. Once the download is complete, click on the button **Apply** to link the Reference Data Set to the workflow we will use later on. You can now close the Data Manager.

Note: the subfolders in the CLC_References/homo_sapiens folder have now a hg19 item each (see figure 3). You will need to use the hg19 sequence as reference in the future steps.

Import the primers

The QIAseq DNA V3 Panel Analysis Plugin includes a new Import tool specifically designed to import the primers specific to a QIAseq DNA V3 Panel.

1. Go to **Import | Import QIAGEN Primers**

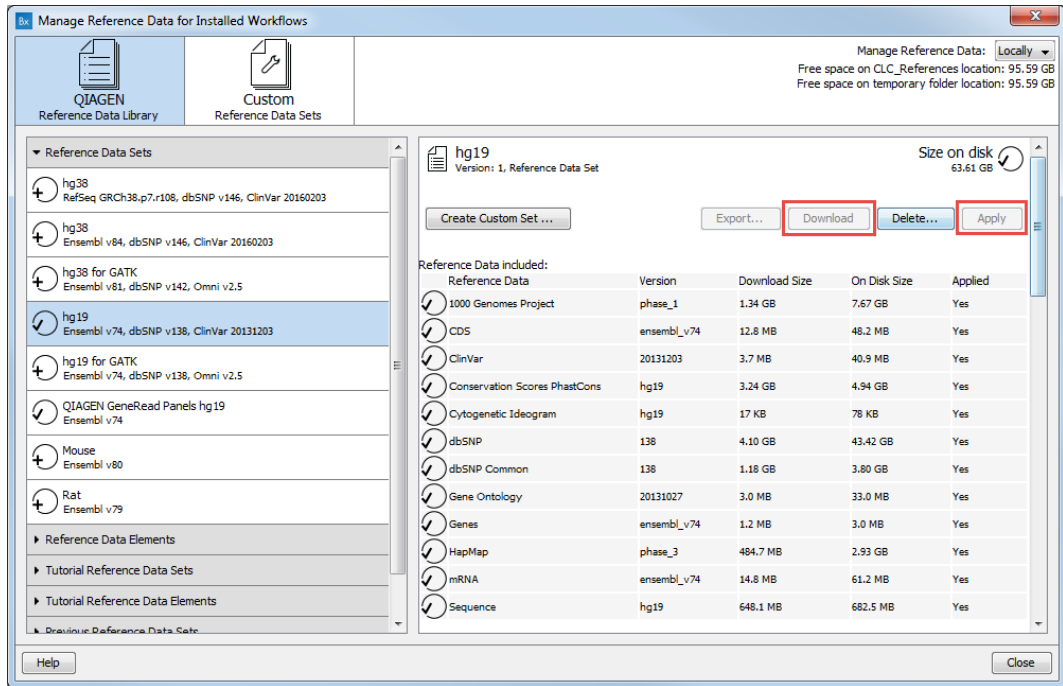


Figure 2: Open the Data Management and download hg19.

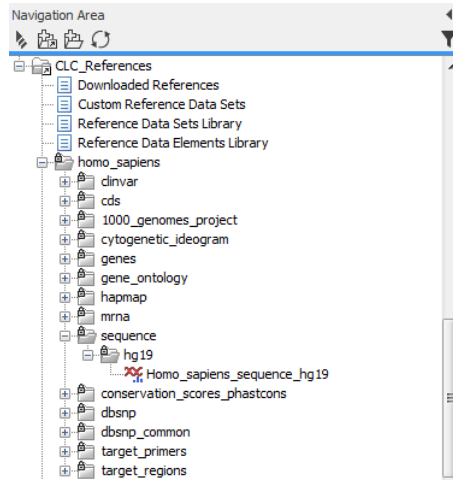


Figure 3: Finding the hg19 reference sequence in CLC_References data folder.

2. In the "Settings" dialog (figure 4), specify the Primer file located on your computer using the folder icon to the right of the field.
3. Specify also a Reference Track in the second field, hg19 in this case. These sequences can be found in the CLC_References located in the Navigation Area after you downloaded the Reference Data Set.
4. Save the primer file in the tutorial folder in the Navigation Area.

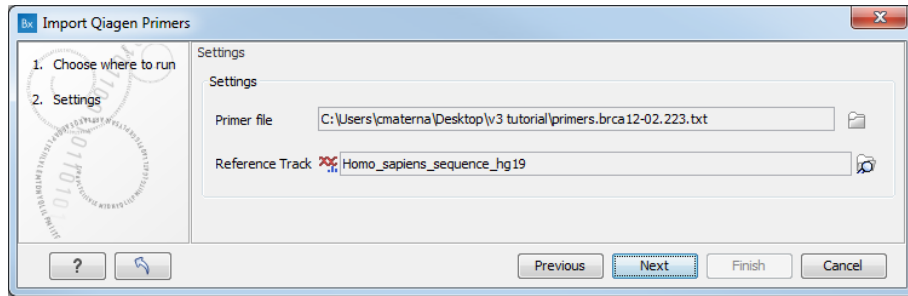


Figure 4: Importing the primers specific to the QIAseq DNA V3 panel used.

Import the Target regions file

The QIAseq DNA Panels primers target different regions of the genome that are specified in a *.bed file. This file is made available to you upon purchase of a kit, but in the case of this tutorial it is included in the data downloaded earlier.

1. Once the file is saved on your computer, click on the **Import** button on the top left of the workbench (1), and use the **Tracks** option (2).
2. In the import wizard (figure 5), choose "BED" from the drop down-menu (3) and select the *.bed file that was provided to you with the QIAseq DNA panel you used (4).

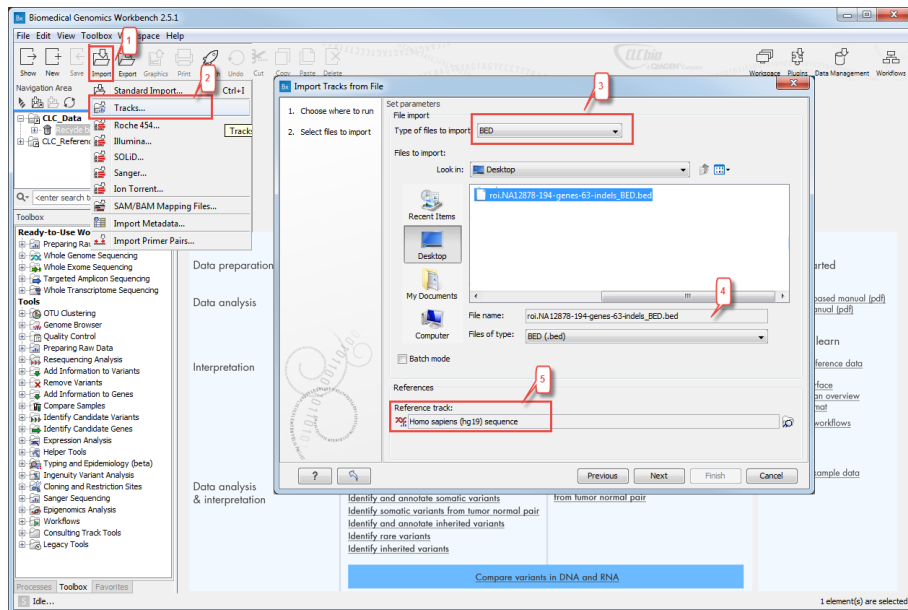


Figure 5: Import the target regions BED file.

3. Specify also a reference sequence (hg19 here) (5).
4. Click **Next** and choose to save the target regions file in the Navigation Area of your workbench.

Import the PCR Adaptor trim list

To import a CLC file, simply choose the **Standard Import** option from the **Import** menu and choose the file called "QIAseq universal 21 bp PCR adaptor trim list". Save it in the folder you created for this tutorial.

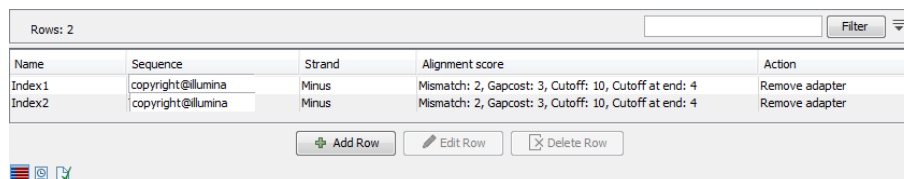
Prepare your reads

As remaining adaptor sequences in the reads might lead to bias in downstream data analysis, we recommend to trim them off.

For this you will have to create a sequencing adaptors list using the oligonucleotide sequences disclosed in the Illumina Customer Sequence Letter found at this address: <http://support.illumina.com/downloads/illumina-customer-sequence-letter.html>.¹

This reads were sequenced using a Illumina Nextera Library Prep Kit.

1. Open the workbench and go to **File | New | Trim Adaptor List (🇺🇸)**
2. Click on the button **Add Row (+)** found at the bottom of the View Area in the New Adaptor Trim List.
3. Name the adapter Index 1. Then copy and paste from the letter the sequence situated to the right of [i7] of the Nextera Index Kit - PCR primer called Index 1 Read. Adjust "Strand" to **Minus** and "Action" to **Remove adaptor**. Leave the other parameters as default (figure 6). Click on the button labeled **Finish** to create the adaptor trim list.



Name	Sequence	Strand	Alignment score	Action
Index1	copyright@illumina	Minus	Mismatch: 2, Gapcost: 3, Cutoff: 10, Cutoff at end: 4	Remove adaptor
Index2	copyright@illumina	Minus	Mismatch: 2, Gapcost: 3, Cutoff: 10, Cutoff at end: 4	Remove adaptor

Figure 6: Creating the adapter list. By default, Mismatch cost is set at 2 and Gap cost at 3. Allow internal matches is checked and its Minimal score set at 10; Allow end matches is also checked and the Minimum score at end is set at 4.

4. Click on the button **Add Row (+)** again.
5. Name the new adapter Index 2. Copy and paste from the letter the sequence situated to the right of the index [i5] of the Nextera Index Kit - PCR primer called Index 2 Read. Adjust "Strand" to **Minus** and "Action" to **Remove adaptor**. Leave the other parameters as default. Click on the button labeled **Finish** to add the adaptor to the list.
6. Go to **File** in the menu bar and choose to **Save** the generated adapter trim list. Name the file **Nextera adapter list** and save to a folder you created for this tutorial in the Navigation Area.
7. Go to **Toolbox | Ready-to-Use Workflows | Preparing Raw Data | Prepare Raw Data**.
8. First select the **reads** (figure 7) and click **Next**.

¹Oligonucleotide sequences © 2007-2013 Illumina, Inc. All rights reserved. No sponsorship or affiliation. Link provided for convenience. QIAGEN not responsible for content at link.

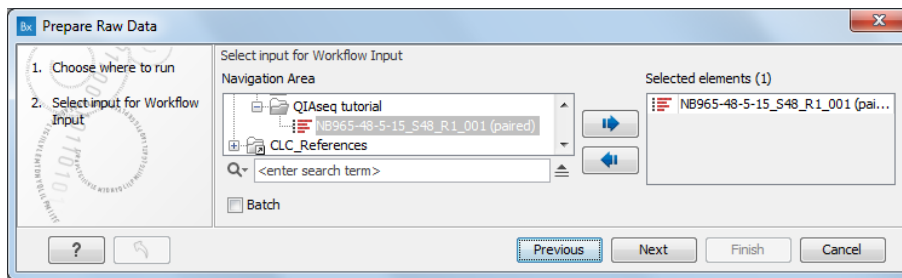


Figure 7: Choose the Illumina paired end read file you imported.

- Then select the **Nextera adapter list** you just created (figure 8). Leave all other parameters as they are set by default and click **Next**.

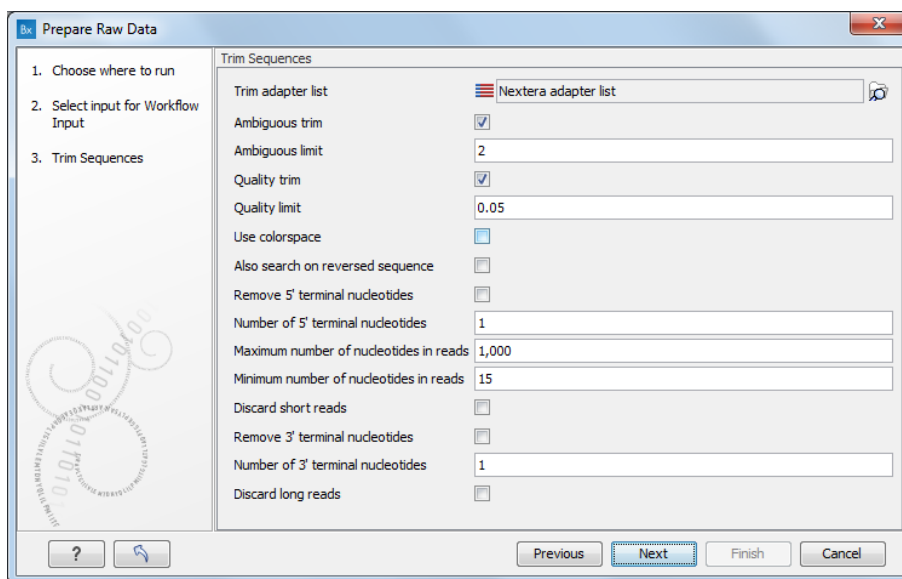


Figure 8: Choose the Nextera adapter list you created previously.

- Choose to save your results in the Navigation Area. Click **Finish** to start the workflow.

The Prepare Raw Data workflow will generate several reports, and up to two outputs of trimmed sequences files per sample. All outputs can be used as input for the Ready-to-Use workflow covered in the next paragraph, but we do not recommend including the broken pairs (orphans) as they may be of sub-optimal quality.

Now we are ready to start the QIAseq DNA V3 Panel Analysis ready-to-use workflow.

How to run the QIAseq DNA V3 Panel Analysis workflow

The QIAseq DNA V3 Panel Analysis workflow can be found in the toolbox under "Targeted Amplicon Sequencing":

Toolbox | Ready-to-Use Workflows | Targeted Amplicon Sequencing (📁) | Somatic Cancer (TAS) (📁) | QIAseq DNA V3 Panel Analysis (📄)

We will run the workflow with the trimmed paired end reads and all the settings set to their default values.

1. Double-click on the **QIAseq DNA V3 Panel Analysis** ready-to-use workflow to run the analysis.
2. Select the sequencing reads that should be analyzed, i.e., the trimmed paired reads that were generated in the previous step (figure 9).

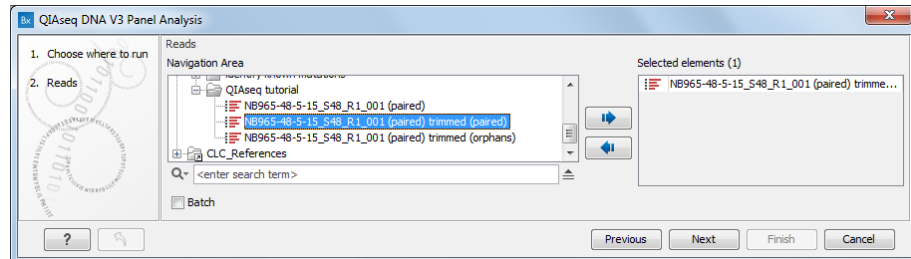


Figure 9: Select the sequencing reads by double-clicking on the file name or by clicking once on the file name and then on the arrow pointing to the right hand side.

3. Select the **QIAseq universal 21 bp PCR adapter trim list** (figure 10).

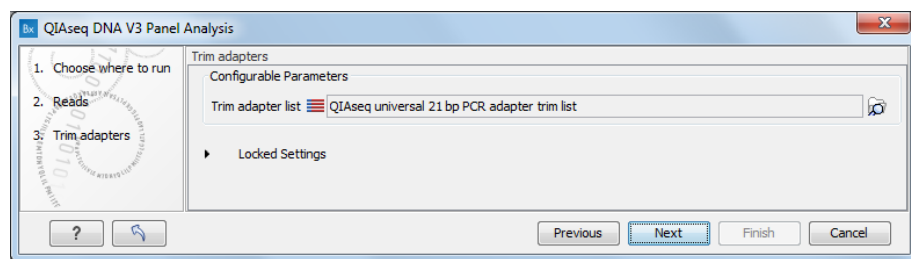


Figure 10: Select the QIAseq universal 21 bp PCR adapter trim list that was provided in the dataset of this tutorial.

4. Select the **primer** track specific to the QIAseq DNA Panel you used to generate the sequenced reads (figure 11).

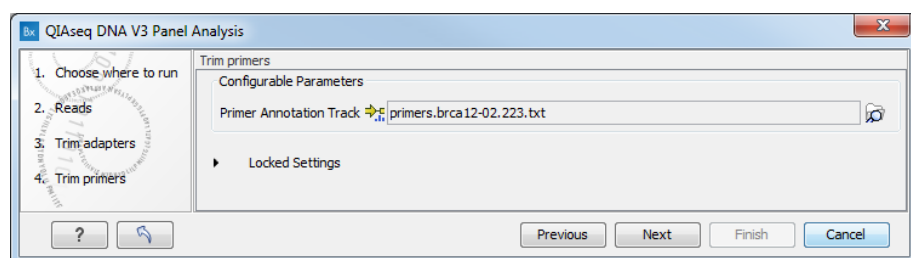


Figure 11: Select the primer file you imported ahead of running the workflow.

5. In the "Low Frequency Variant detection" wizard window (figure 12), specify the **target region BED** file you imported previously.
6. The two following wizard windows correspond to variant filters that have been optimized to suit QIAseq Targeted DNA panels and provide true variants while minimizing false positives and false negatives. Therefore we recommend the use of default values set for "Filter on Variant Annotations" (in particular, do not load any annotations, nor try to change the name of the filters in the first column, as it would disable the filter completely) and "Filter on Variant Allele Frequency".

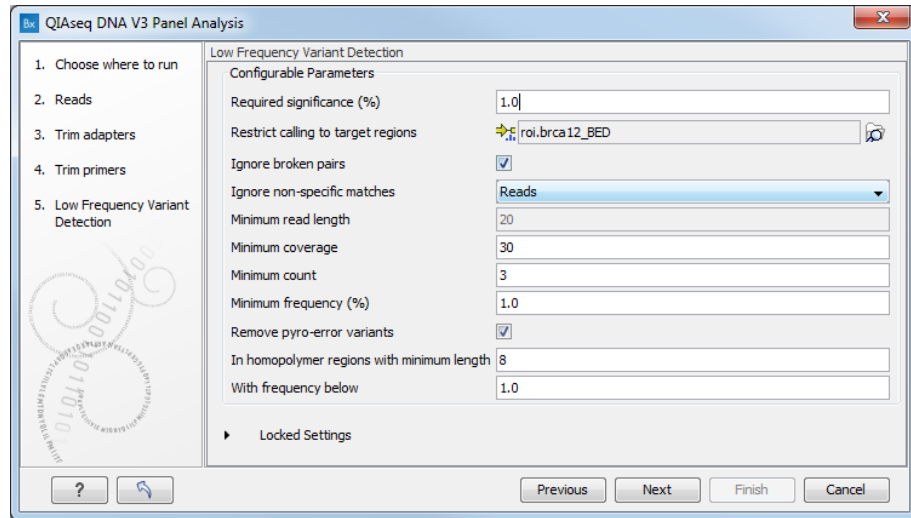


Figure 12: Select the target regions file you imported ahead of running the workflow.

7. Finally, in the last wizard step, choose to **Save** the results of the workflow and specify a location in the Navigation Area before clicking **Finish**.

Output from the QIAseq DNA V3 Panel Analysis

The QIAseq DNA V3 Panel Analysis workflow produces the following outputs:

- a Variant track (📊)
- a Genome Browser View (📄)
- a Unique Molecular Index Groups Report (📄)
- a read mapping of the UMI Reads (📄)

The report and read mapping are available in a subfolder as seen in figure 13.

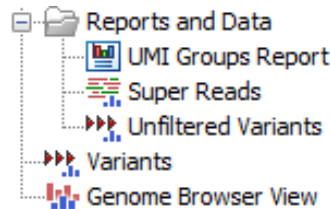


Figure 13: Output from the QIAseq DNA V3 Panel Analysis workflow.

Double click on the Genome Browser View. The file opens in a split screen with the Browser View and the corresponding variant table. In the table, filter for variants for which the allele is not identical to the reference: open the advanced filter by clicking on the arrow in the top right corner of the table. Fill in the filtering fields with "**Reference allele**", "=", and type "**No**".

QIAseq panels allow the detection of very low frequency variants. To observe this is in the tutorial data, sort the "**Frequency**" column in the variant table in ascending order by clicking on the

Tutorial

header of that column. Variants with a frequency ("Count" divided by "Coverage") as low as 2 appear at the top of the table. Clicking on a particular row in the variant table causes the Genome Browser to zoom in to the nucleotide level with the focus on that variant position. This allows you to view the coverage and the context of this variant, e.g., Super Reads, mRNA annotations, CDS annotations, and so on.
