



Tutorial

Find Very Low Frequency Variants With QIAGEN GeneRead Panels

February 3, 2022

— Sample to Insight —

Find Very Low Frequency Variants With QIAGEN GeneRead Panels

This tutorial uses the capabilities of the CLC Genomics Workbench and the Biomedical Genomics Analysis plugin to find very low frequency variants in Targeted Amplicon sequences generated using a QIAGEN GeneRead DNaseq panel kit (V2).


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

- Import Illumina paired reads in the workbench.
- Import an adapter list and use the Prepare Raw Data Template Workflow to remove remaining adapters from the reads and generate QC reports.
- Find low frequency variants with the QIAGEN GeneRead Panel Analysis Template Workflow, steps include:
 - Alignment of data to reference
 - Removal of primers and primer-dimer artifacts
 - Very Low Frequency Variant Detection
 - Generation of additional reports (alignment report, coverage report)
 - Annotation of variants
 - Setting up a Track List for visualization of the results

Prerequisites For this tutorial, you must have installed the Biomedical Genomics Analysis plugin. How to install plugins is described here: <http://resources.qiagenbioinformatics.com/manuals/clcgenomicsworkbench/current/index.php?manual=Install.html>.

Import data and prepare your reads

This tutorial makes use of Targeted Amplicon Sequencing data provided by Dr Matthew Smith of University Hospitals Birmingham NHS Foundation Trust using HDx Reference Standards from Horizon's diagnostics division. The sample used comes from a cell line in which variants were introduced at very low frequency, and was sequenced on an Illumina platform. 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/QIAGEN_GeneRead_panels.zip.
2. Start the workbench.
3. Import the reads via the toolbar: **Import**  | **Illumina**
 - Select the two fastq paired files.
 - Under "Import files and options" section (figure 1), ensure that:
 - **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 **Next**.
4. Click **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**.

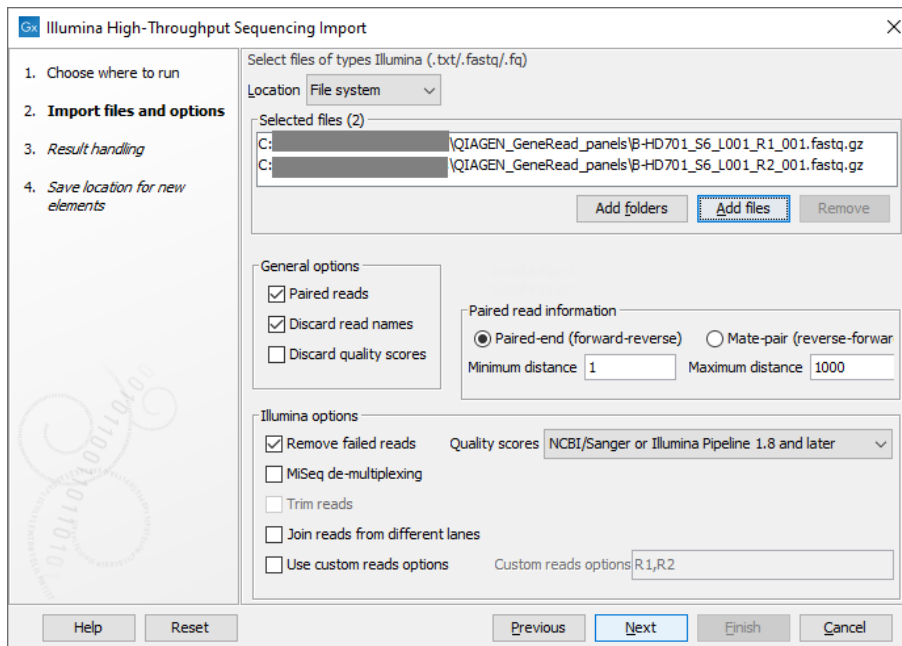


Figure 1: Import data using a specified importer.

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

For this you will have to use an adapter list. In this tutorial we will use Illumina TruSeq Exome Adapter list located in the already downloaded sample data as a CLC object.

Import the adapter list via the toolbar: **Import** (📁) | **Standard Import**

1. Select the "Exome_TrueSeq_read_through_adapter_trim_list" file from the downloaded and unzipped sample data and leave the Automatic import checked as shown in figure 2.

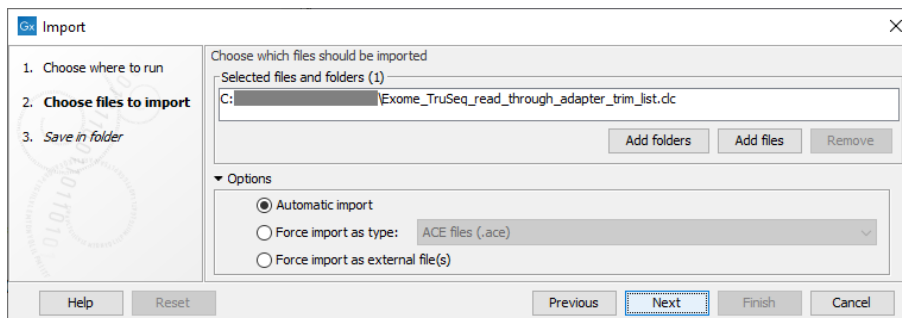


Figure 2: Import an adapter list to be used to trim off adapters still remaining on the reads.

2. Click **Save** in the wizard page that appears, choose the folder you wish to save the adapter list to and click **Finish**.

Now we are ready to trim the adapters from our reads. Go to

Toolbox | Template Workflows | Biomedical Workflows  | **Preparing Raw Data**  | **Prepare Raw Data** 

1. This opens a dialog where you select the HD701 sample as shown in figure 3.

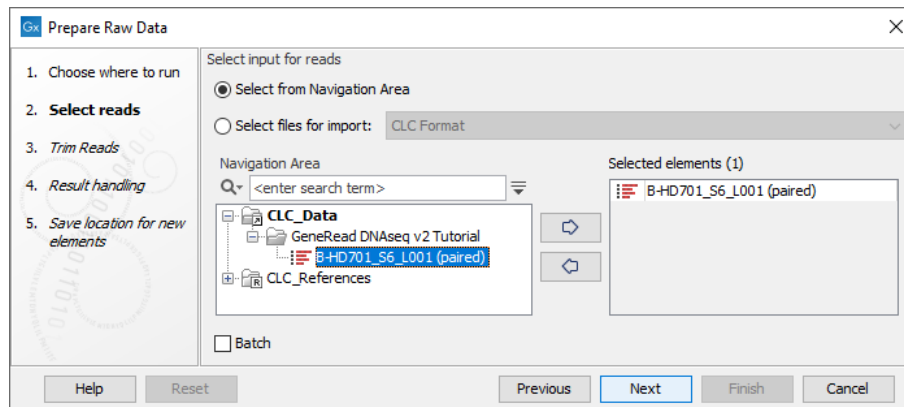


Figure 3: Select the raw sequencing reads by double-clicking on the file name or by clicking once on the file name and then on the arrow pointing to the right hand side.

2. In the next dialog (figure 4). Click on the browse icon next to the "Trim Adapter list" field, and in the dialog that opens, select the "Exome_TrueSeq_read_through_adapter_trim_list" that you have just imported and click OK. Leave all parameters as they are set by default and click **Next**.

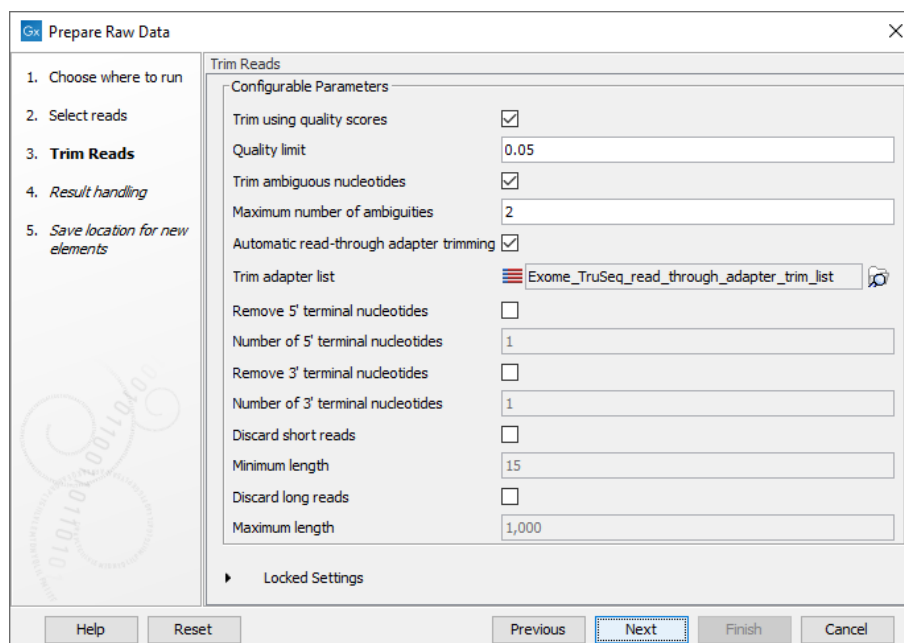


Figure 4: Choosing to trim for adapter sequence using a adapter list.

3. In the Result handling window, choose to **Save** the outputs files, and click on the button labeled **Finish**.

Once the analysis is complete, three reports and two trimmed sequence lists (📄) have been generated (figure 5).

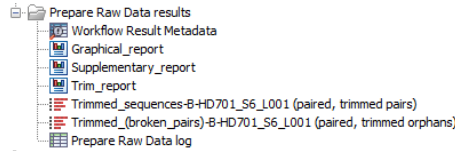


Figure 5: Output from the "Prepare Raw Data" workflow.

Open the Trim report (figure 6). It is meant to be used as a quality check, mainly to see that the adapter trimming worked as expected. Our reads are ready to be used for variant detection.

4 Detailed trim results

Trim	Input reads	No trim	Trimmed	Nothing left or Discarded
Trim on quality	1,304,138	1,284,224	19,914	0
Ambiguity trim	1,304,138	1,304,131	7	0
Adapter trimming	1,304,138	762,512	541,414	212

Figure 6: Results of each of the selected trimming parameters.

Running the QIAGEN GeneRead Panel Analysis workflow

The **QIAGEN GeneRead Panel Analysis** workflow is an automated way to run a series of tools using the output of one as input for the next. The first step in the template workflow is mapping of the sequencing reads to the human reference sequence. This is followed by a local realignment step, which is included to improve the variant detection that follows directly after a primer trimming step. After variant detection, the variants are annotated with gene names, exon numbers, amino acid changes, conservation scores, information from relevant variants present in the ClinVar database, and information from common variants present in the common dbSNP, HapMap, and 1000 Genomes database. Furthermore, a detailed target regions mapping report is created that allows inspection of the coverage and mapping specificity in the target regions.

To run this workflow, go to:

Toolbox | Template Workflows | Biomedical Workflows (🧬) | Targeted Amplicon Sequencing (📄) | Somatic Cancer (🔴) | QIAGEN GeneRead Panel Analysis (🌐)

1. Select the sequencing reads that should be analyzed, i.e., the trimmed ones you generated previously (figure 7).

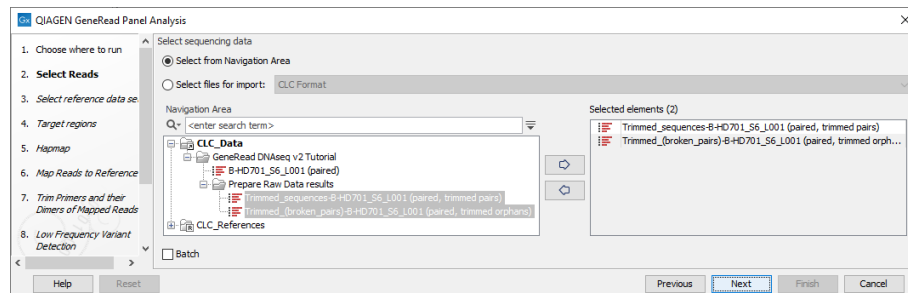


Figure 7: Select the sequencing reads by double-clicking on the file name or by clicking once on the file name and then on the arrow pointing to the right hand side.

- In the next dialog, select the QIAGEN GeneRead Panels hg19 set as seen in figure 8. Download to Workbench if not there already. When download is complete click **Next**.

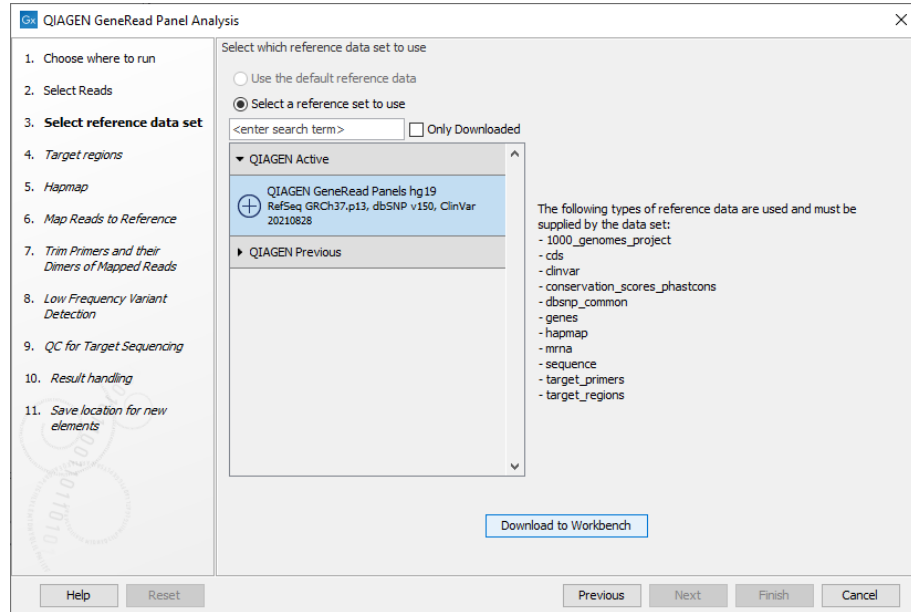


Figure 8: The relevant Reference Data Set is highlighted. In the text to the right, the types of reference data needed are listed.

- In the next dialog select the **NGHS-201X_Cancer_Actionable_Mutations** target regions from the drop down menu as seen in figure 9. Click **next**.

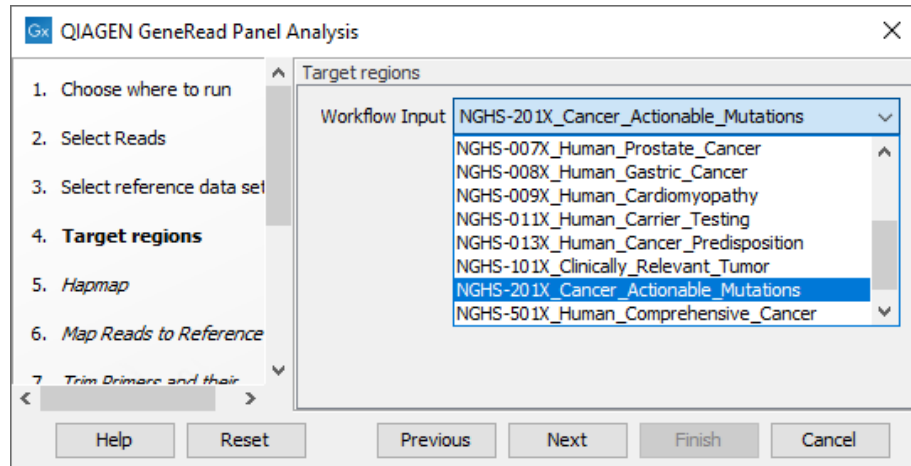


Figure 9: The target regions matching the panel is selected.

- All possible Hapmap populations are pre-selected. You can choose to deselect the panels/populations you are not interested in but we will keep them all selected in the case of this tutorial, so click **Next**.
- In the "Map Reads to reference" step (figure 10), leave the read mapper configured as it is by default, i.e., the "Cost of insertions and deletions" is set to "Affine gap cost". Click **Next**.

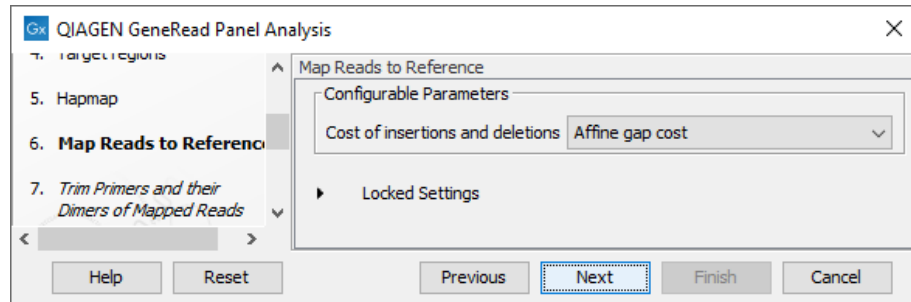


Figure 10: Leave the "Cost of insertions and deletions" to "Affine gap cost" (default).

- In the next dialog (figure 11), the target primers for primer trimming can be specified from the drop down menu. Select again the **NGHS-201X_Cancer_Actionable_Mutations** and leave the parameter "Only keep reads that have hit a primer" enabled (it is already checked by default).

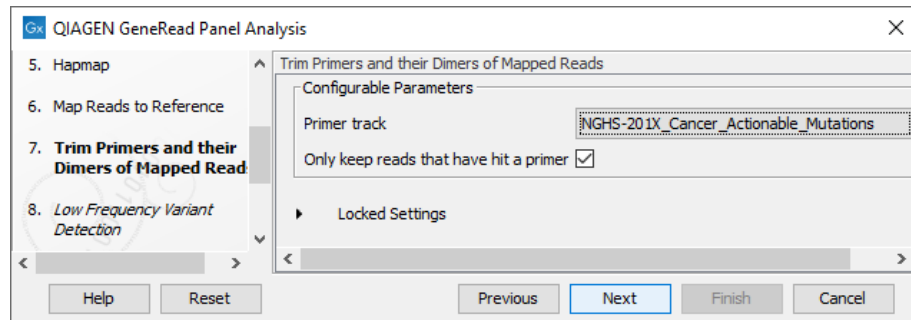



Figure 11: Select the primer track from the drop-down list.

- In the next wizard step (figure 12) you can specify the parameters for variant detection. As we wish to detect very low frequency variant, set the **Minimum frequency** to **0.5%**. And to avoid having too many false positive, we will raise the **Minimum coverage** to **500**.
- In the "QC for Targeted Sequencing" step, set the Minimum coverage to 500 and leave the other parameters as default before clicking **Next**.
- Choose to save the results, create a new folder called results within your tutorial folder, and click on the button labeled **Finish**.

Output from the QIAGEN generead panel workflow

Open the Track List called Genome Browser View () and focus for now on the variant table that opens automatically in split view to get an overview of the identified variants (see 14).

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 "No". Only 36 variants are now left in the table. To find the very low frequency variants, click on the header "Frequency". You can now easily check coverage for each of these variants by clicking on a table row. The Track List will zoom in to the chosen variant at the nucleotide level, allowing you to see the variant in the context of the mapped reads, an amino acid track, and various databases.

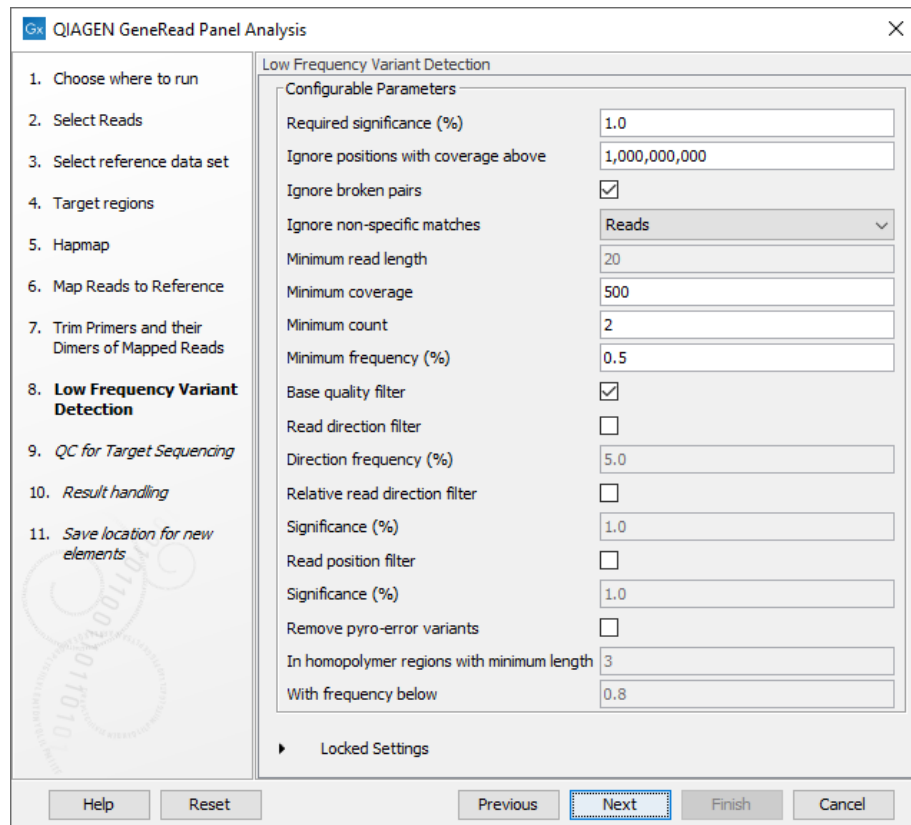


Figure 12: In this wizard step the parameters for variant detection can be adjusted.

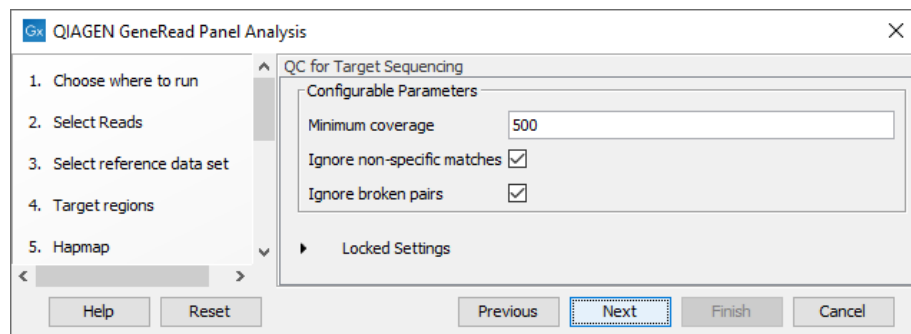


Figure 13: Set the minimum coverage to 500.

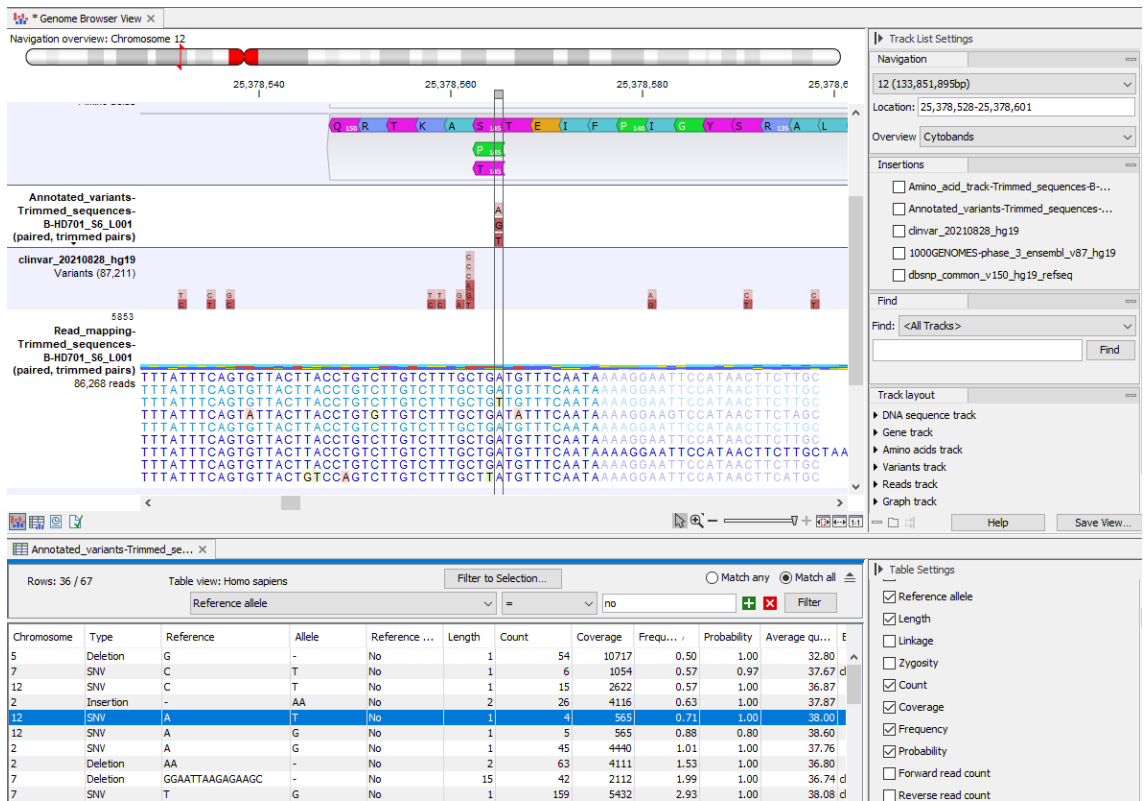


Figure 14: Filter and order the variant table to find the very low frequency variants.