Tutorial

Immune Repertoire Analysis

December 3, 2025

QIAGEN Aarhus A/S · Kalkværksvej 5, 11. · DK - 8000 Aarhus C · Denmark
digitalinsights.qiagen.com · ts-bioinformatics@qiagen.com

Sample to Insight

Immune Repertoire Analysis

The purpose of this tutorial is to demonstrate how *CLC Workbench* and the *Biomedical Genomics Analysis* plugin can be used to detect and analyze clonotypes.

We focus on the following:

- Import data.
- Detect clontypes using a template workflow.
- Interpret the results.
- Compare the detected clonotypes in two samples.

Data used in this tutorial

This tutorial uses data produced with a QIAseq Targeted RNA-seq Panel for T-cell Receptors from a sample sequenced twice on an Illumina platform using 300 and 500 cycles. The 300-cycle run produces shorter R2 reads, which leads to a smaller portion of the V segment being covered and therefore increases the number of clonotypes with ambiguous V segment assignments.

To complete the tutorial in a reasonable amount of time, the data used is a subsample of the full sequencing reads.

Prerequisites

For this tutorial, you must be working with *CLC Genomics Workbench* 26 and *Biomedical Genomics Analysis* plugin 26 or higher. Note that higher versions may produce slightly different results than those shown here.

Installing plugins is described in the CLC Genomics Workbench manual.

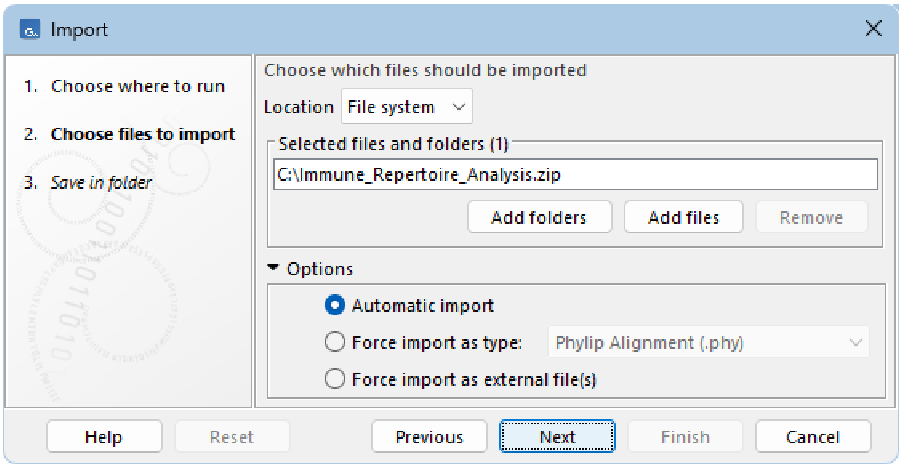
General tips

- Throughout this tutorial, we provide links to relevant manual pages, which we recommend exploring for additional details.
- Tools and workflows can be found in the Toolbox, but it is often easier to launch them using Quick Launch (🚀), found in the top toolbar (shortcut Ctrl+Shift+T or ⌘ +Shift+T on Mac). Quick Launch displays the full Toolbox path, making it easy to identify the location of the tool or workflow if needed.
- The in-built manual can be accessed by clicking the **Help** button on wizards or by selecting the **Help** option under the **Help** menu.
- Within wizards, the **Reset** button can be used to change settings to their default values.
- Columns in tables can be hidden by unchecking their name in the Side Panel.
- Columns in tables can be used to sort the rows, by successively clicking on the column name until the desired order (indicated by an arrow next to the column name) is achieved.
- Colors and gradients in plots can be changed by clicking on them in the Side Panel.
- Most of the tools of *CLC Genomics Workbench* require multiple inputs. When many data elements need to be selected, all elements located under a folder can be added by using the options **Add folder contents** or **Add folder contents (recursively)** found in the right-click menu.
- Many data elements produced by *CLC Genomics Workbench* tools have multiple views, indicated as icons in the lower left corner of elements opened in the View Area. Clicking on one of the view icons while pressing the Ctrl (⌘ on Mac) key will open in split view such that both views are visible at the same time. Often, if viewing a table and a graphical representation in split view, selecting entries in the table will highlight them in the graphical representation. The order of the views can be changed using drag and drop, see Arrange views in View Area.

Import the data

We start by downloading and importing the tutorial data.

1. Download the tutorial data.
2. Start the *CLC Genomics Workbench*.
3. Import the data using Standard Import:
   (a) Launch **Standard Import** () using Quick Launch ().
   (b) Locate the tutorial data using the **Add files** button and select **Automatic import** (figure 1).



Import

1. Choose where to run
2. Choose files to import
3. Save in folder

Choose which files should be imported

Location File system

Selected files and folders (1)

C:\Immune_Repertoire_Analysis.zip

Options

● Automatic import

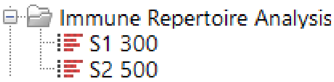
○ Force import as type: Phylip Alignment (.phy)

○ Force import as external file(s)

Figure 1: *Standard Import configured to import the tutorial data.*

   (c) In the next step, select a suitable location in the Navigation Area to save the imported data and click on **Finish**.

Once the import is completed, the folders and data elements are visible in the Navigation Area (figure 2).



Immune Repertoire Analysis
S1 300
S2 500

Figure 2: *The imported tutorial data in the Navigation Area.*

Detect clonotypes

We will now use the Analyze QIAseq Targeted TCR (Illumina) template workflow to analyze the tutorial data: identify T cell receptor (TCR) sequences and characterize the T cell clonotypes repertoire. This workflow has been designed for data generated using a QIAseq Targeted RNA-seq Panel for T-cell Receptor panel. If you run this workflow on your own data, please note that template workflows are provided as example workflows and may need to be customized to meet the specific requirements of your data.

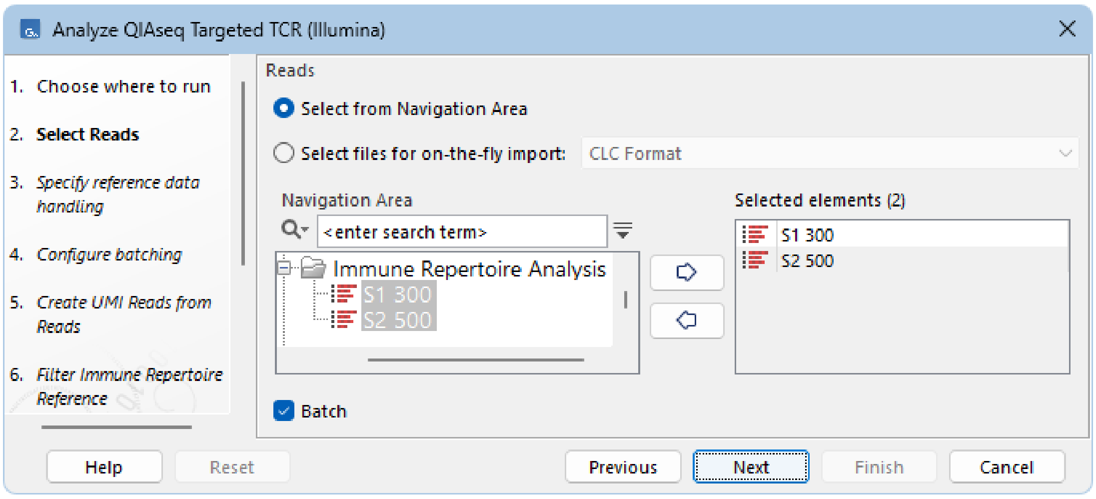
To see the contents of the workflow, locate it in the Toolbox:

Template Workflows | **Biomedical Workflows** () | **QIAseq Sample Analysis** () | **Other QIAseq Workflows** () | **Analyze QIAseq Targeted TCR (Illumina)** ()

right-click on its name and choose **Open Copy of Workflow**.

We will now run the workflow:

1. Launch the workflow using Quick Launch (🚀) or by double-clicking its name in the Toolbox.
2. In the first wizard step, "Select Reads", specify the data to be analyzed by selecting the imported reads. Make sure to check the **Batch** option below the data selection area (figure 3).



Analyze QIAseq Targeted TCR (Illumina)

1. Choose where to run
2. **Select Reads**
3. *Specify reference data handling*
4. *Configure batching*
5. *Create UMI Reads from Reads*
6. *Filter Immune Repertoire Reference*

Reads

● Select from Navigation Area

○ Select files for on-the-fly import: CLC Format

Navigation Area

Immune Repertoire Analysis
- S1 300
- S2 500

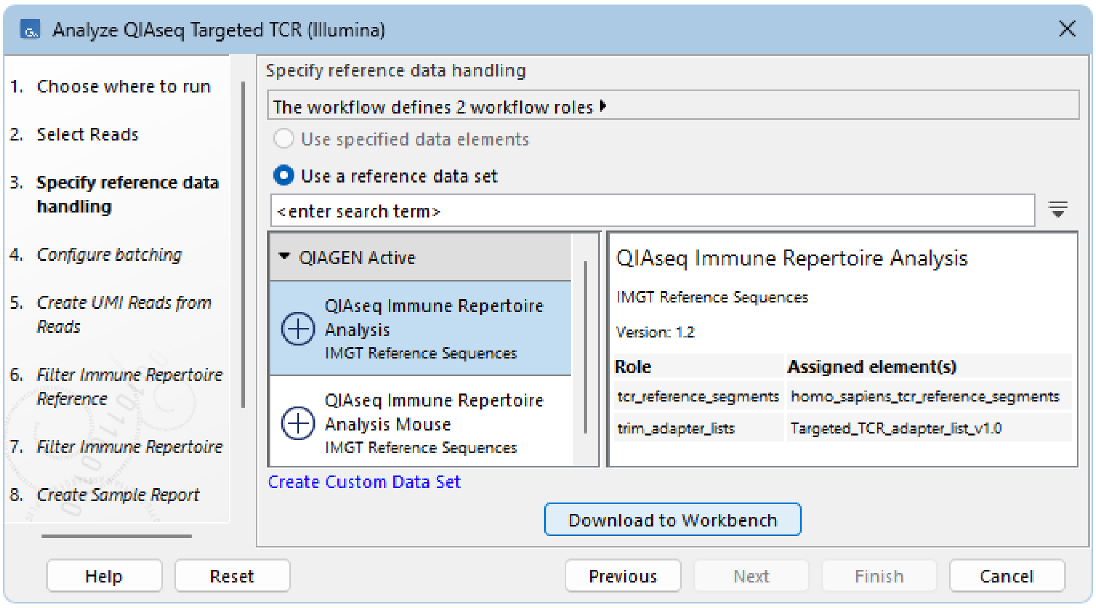
Selected elements (2)

S1 300
S2 500

☑ Batch

Figure 3: *The two samples are used as input and the Batch option is checked.*

3. In the next step, "Specify reference data handling", select the "QIAseq Immune Repertoire Analysis" Reference Data Set (figure 4). Click **Download to Workbench** if the data set has not already been downloaded. Then click **Next**.



Analyze QIAseq Targeted TCR (Illumina)

1. Choose where to run
2. Select Reads
3. **Specify reference data handling**
4. *Configure batching*
5. *Create UMI Reads from Reads*
6. *Filter Immune Repertoire Reference*
7. *Filter Immune Repertoire*
8. *Create Sample Report*

Specify reference data handling

The workflow defines 2 workflow roles ▸

○ Use specified data elements

● Use a reference data set

▾ QIAGEN Active

QIAseq Immune Repertoire Analysis
IMGT Reference Sequences

QIAseq Immune Repertoire Analysis Mouse
IMGT Reference Sequences

QIAseq Immune Repertoire Analysis

IMGT Reference Sequences

Version: 1.2

Role	Assigned element(s)
tcr_reference_segments	homo_sapiens_tcr_reference_segments
trim_adapter_lists	Targeted_TCR_adapter_list_v1.0

Figure 4: *Reference data to be downloaded*

4. In the next wizard steps, keep the default settings.

5. In the last step, make a new subfolder in "Immune Repertoire Analysis" called "Results" and chose to save the workflow results there. Make sure to check the **Create subfolders per batch unit** option.

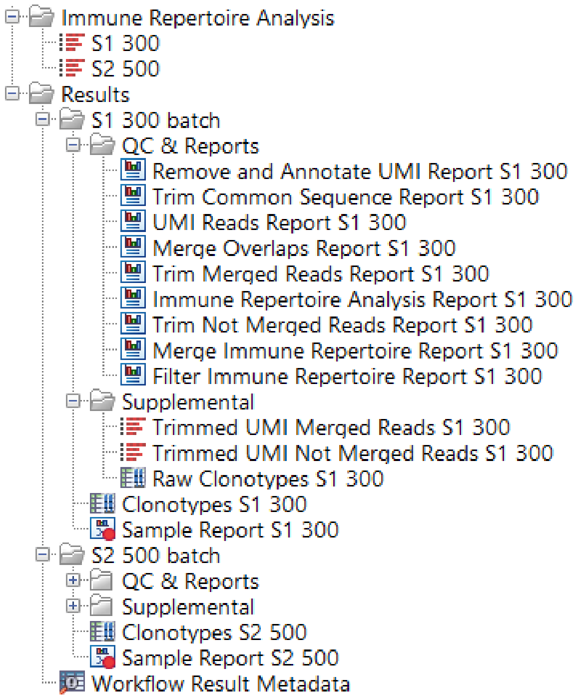
   Click on **Finish**.

   The workflow will now execute. The progress can be monitored under the Processes tab in the Toolbox.

For this data, all chains have been sequenced. For panels targeting only specific chains, choose the relevant ones in the **Filter Immune Repertoire Reference** wizard step.

Interpret the results

Results from the workflow are placed in the "Results" folder (figure 5).



- Immune Repertoire Analysis
  - S1 300
  - S2 500
- Results
  - S1 300 batch
    - QC & Reports
      - Remove and Annotate UMI Report S1 300
      - Trim Common Sequence Report S1 300
      - UMI Reads Report S1 300
      - Merge Overlaps Report S1 300
      - Trim Merged Reads Report S1 300
      - Immune Repertoire Analysis Report S1 300
      - Trim Not Merged Reads Report S1 300
      - Merge Immune Repertoire Report S1 300
      - Filter Immune Repertoire Report S1 300
    - Supplemental
      - Trimmed UMI Merged Reads S1 300
      - Trimmed UMI Not Merged Reads S1 300
      - Raw Clonotypes S1 300
    - Clonotypes S1 300
    - Sample Report S1 300
  - S2 500 batch
    - QC & Reports
    - Supplemental
    - Clonotypes S2 500
    - Sample Report S2 500
  - Workflow Result Metadata

Figure 5: *The "Results" folder in the Navigation Area.*

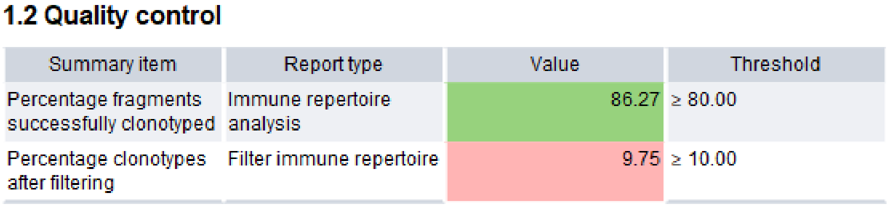
For each sample, the folder contains the following contents, among others:

- A **QC & Reports** subfolder containing all reports produced by the workflow.
- A **Supplemental** subfolder containing partially processed data that can be used for re-running parts of the analysis using different settings, for example when merging or filtering the clonotypes.
- A Clonotypes element containing the detected TCR clonotypes.
- A Sample Report summarizing information from all the reports located in the QC & Reports folder.

Quality control

It is important to first verify that the data quality is satisfactory.

Open the **Sample Report S1 300**, found in the "Results/S1 300 batch" folder. The Quality control section contains different summary items that can be used to assess the quality of the reads (figure 6). These summary items can be configured in the "Create Sample Report" wizard step when launching the workflow.



1.2 Quality control

Summary item	Report type	Value	Threshold
Percentage fragments successfully clonotyped	Immune repertoire analysis	86.27	≥ 80.00
Percentage clonotypes after filtering	Filter immune repertoire	9.75	≥ 10.00

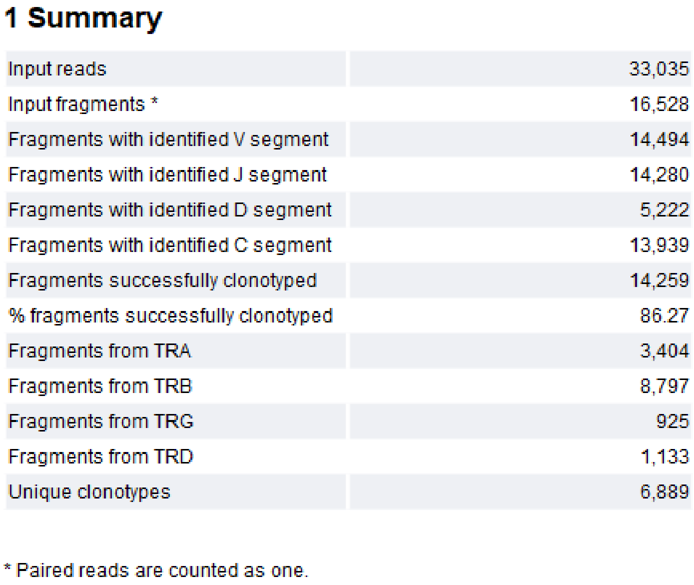
Figure 6: *The sample report contains summary items for assessing the quality of the reads.*

As the tutorial data is a subsample of the full sequencing reads, a relatively low percentage of clonotypes has a Minimum count of at least 2, and the "Percentage clonotype after filtering" is marked in red. This does not affect the analysis here.

Review the reports

To get a quick overview of the identified, merged, and filtered clonotypes, open the following reports in the **QC & Reports** folder, which provide a **Summary** section followed by various statistics and properties of the sample:

- Immune Repertoire Analysis Report. The **Summary** section (figure 7) provides counts of input fragments, fragments with identified segments for each segment type, and clonotyped fragments for each chain type. In most tissues of both human and mouse, TRA/TRB T cells are far more common than TRG/TRD cells, so the number of fragments for these chains is expected to be substantially higher.



1 Summary

Input reads	33,035
Input fragments *	16,528
Fragments with identified V segment	14,494
Fragments with identified J segment	14,280
Fragments with identified D segment	5,222
Fragments with identified C segment	13,939
Fragments successfully clonotyped	14,259
% fragments successfully clonotyped	86.27
Fragments from TRA	3,404
Fragments from TRB	8,797
Fragments from TRG	925
Fragments from TRD	1,133
Unique clonotypes	6,889

* Paired reads are counted as one.

Figure 7: *Summary section for the "Immune Repertoire Analysis Report".*

The rest of the report refers to the initially identified clonotypes: **Raw Clonotypes** found in the **Supplemental** folder.

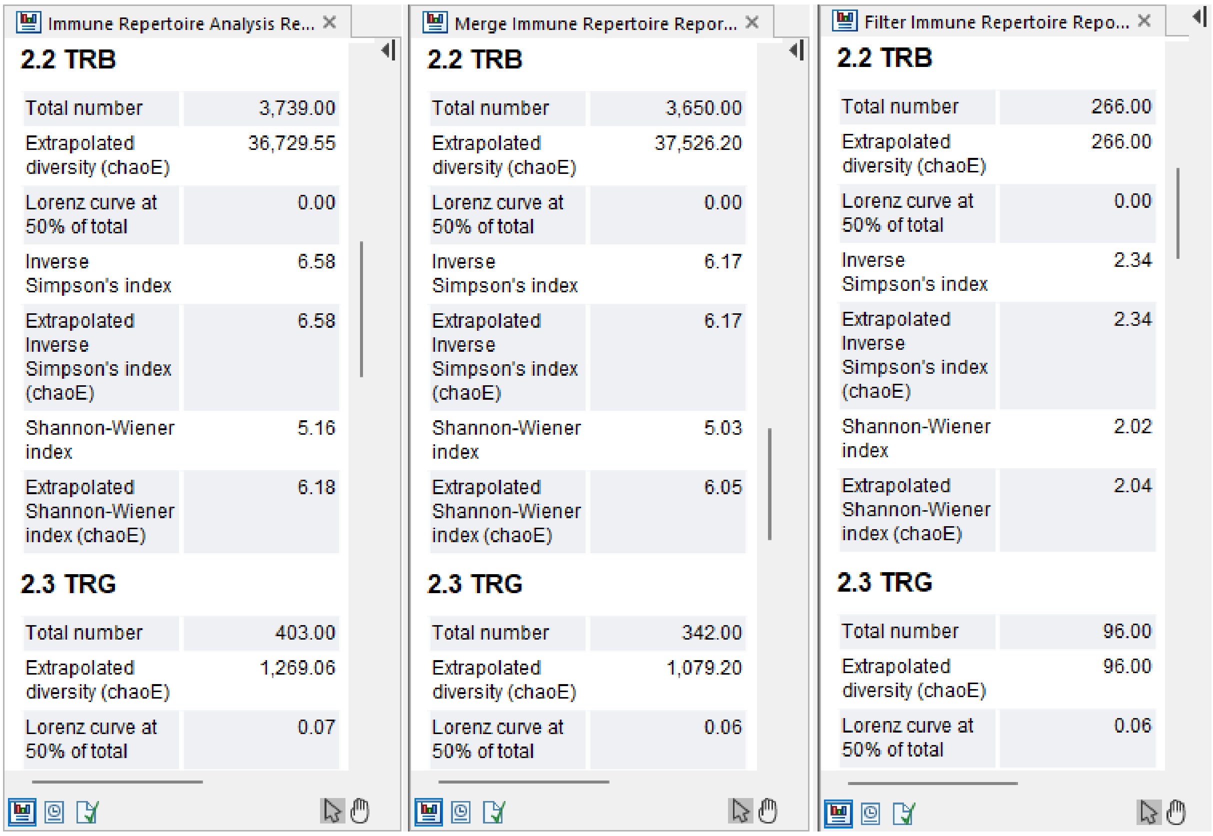
- Merge Immune Repertoire Report. The **Summary** section provides a summary of the performed merging.

  The rest of the report refers to the clonotypes after merging.

- Filter Immune Repertoire Report. The **Summary** section provides a summary of how many clonotypes have been filtered.

  The rest of the report refers to the final detected TCR clonotypes following filtering: **Clonotypes**.

The **Diversity indices** section shows the diversity of a sample using multiple metrics. Low values indicate low diversity, with a few dominating clonotypes, whereas high values indicate greater diversity, reflecting a repertoire with many different clonotypes present at modest frequencies (figure 8). As clonotypes are first merged and then filtered, the number of unique clonotypes decreases at each step, leading to a reduction in diversity. The higher prevalence of TRA/TRB T cells compared with TRG/TRD cells is also apparent in the diversity metrics.



Immune Repertoire Analysis Re...

2.2 TRB

Total number	3,739.00
Extrapolated diversity (chaoE)	36,729.55
Lorenz curve at 50% of total	0.00
Inverse Simpson's index	6.58
Extrapolated Inverse Simpson's index (chaoE)	6.58
Shannon-Wiener index	5.16
Extrapolated Shannon-Wiener index (chaoE)	6.18

2.3 TRG

Total number	403.00
Extrapolated diversity (chaoE)	1,269.06
Lorenz curve at 50% of total	0.07

Merge Immune Repertoire Repor...

2.2 TRB

Total number	3,650.00
Extrapolated diversity (chaoE)	37,526.20
Lorenz curve at 50% of total	0.00
Inverse Simpson's index	6.17
Extrapolated Inverse Simpson's index (chaoE)	6.17
Shannon-Wiener index	5.03
Extrapolated Shannon-Wiener index (chaoE)	6.05

2.3 TRG

Total number	342.00
Extrapolated diversity (chaoE)	1,079.20
Lorenz curve at 50% of total	0.06

Filter Immune Repertoire Repo...

2.2 TRB

Total number	266.00
Extrapolated diversity (chaoE)	266.00
Lorenz curve at 50% of total	0.00
Inverse Simpson's index	2.34
Extrapolated Inverse Simpson's index (chaoE)	2.34
Shannon-Wiener index	2.02
Extrapolated Shannon-Wiener index (chaoE)	2.04

2.3 TRG

Total number	96.00
Extrapolated diversity (chaoE)	96.00
Lorenz curve at 50% of total	0.06

Figure 8: *Diversity indices of TRB and TRG chains at each processing step: initial identification, merging, and filtering.*

The **Rarefaction** section contains rarefaction curves for each chain, illustrating how many unique clonotypes are expected as a function of the repertoire's diversity (figure 9). When only a few clonotypes have been detected, there is a high probability that the next read will represent a new clonotype, causing the curve to rise steeply. As more clonotypes are detected, it becomes increasingly likely that subsequent reads represent clonotypes already observed, so the curve gradually flattens. Because the rarefaction extrapolation relies on the number of clonotypes with a count of 1, the most representative view of the repertoire is obtained from the **Merge Immune Repertoire Report**, before clonotypes are filtered. As the tutorial data is a subsample of the full sequencing reads, the rarefaction curve does not reach its plateau, indicating that many more clonotypes would be discovered if the full data were analyzed.

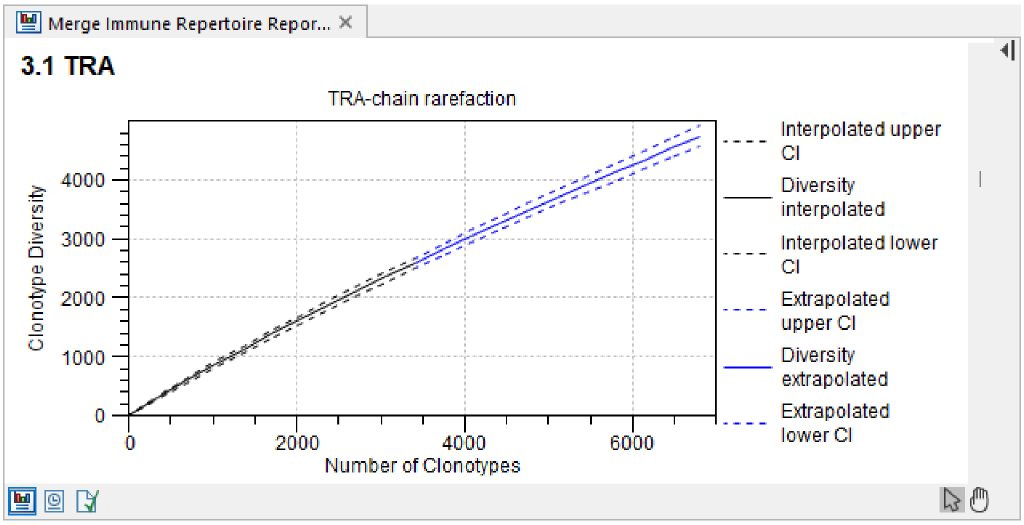


Figure 9: *Rarefaction curve for the TRA chain from "Merge Immune Repertoire Report".*

The **Cumulative frequencies of clonotypes** section shows the cumulative frequencies of clonotypes for each chain, with clonotypes ranked by descending count (figure 10). These curves indicate whether the repertoire is dominated by a few clonotypes that account for most of the reads, resulting in a steep initial rise followed by a plateau, or whether it contains many clonotypes with more evenly distributed frequencies, producing a more linear curve. For the tutorial data, the cumulative frequencies curves are more linear, which may reflect both the inherent repertoire diversity and the fact that the tutorial data is a subsample of the full sequencing reads.

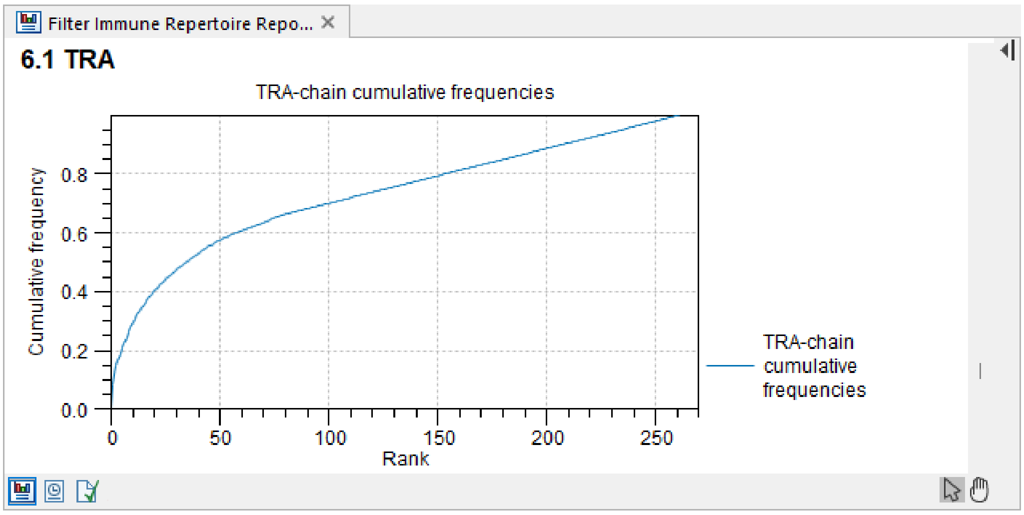
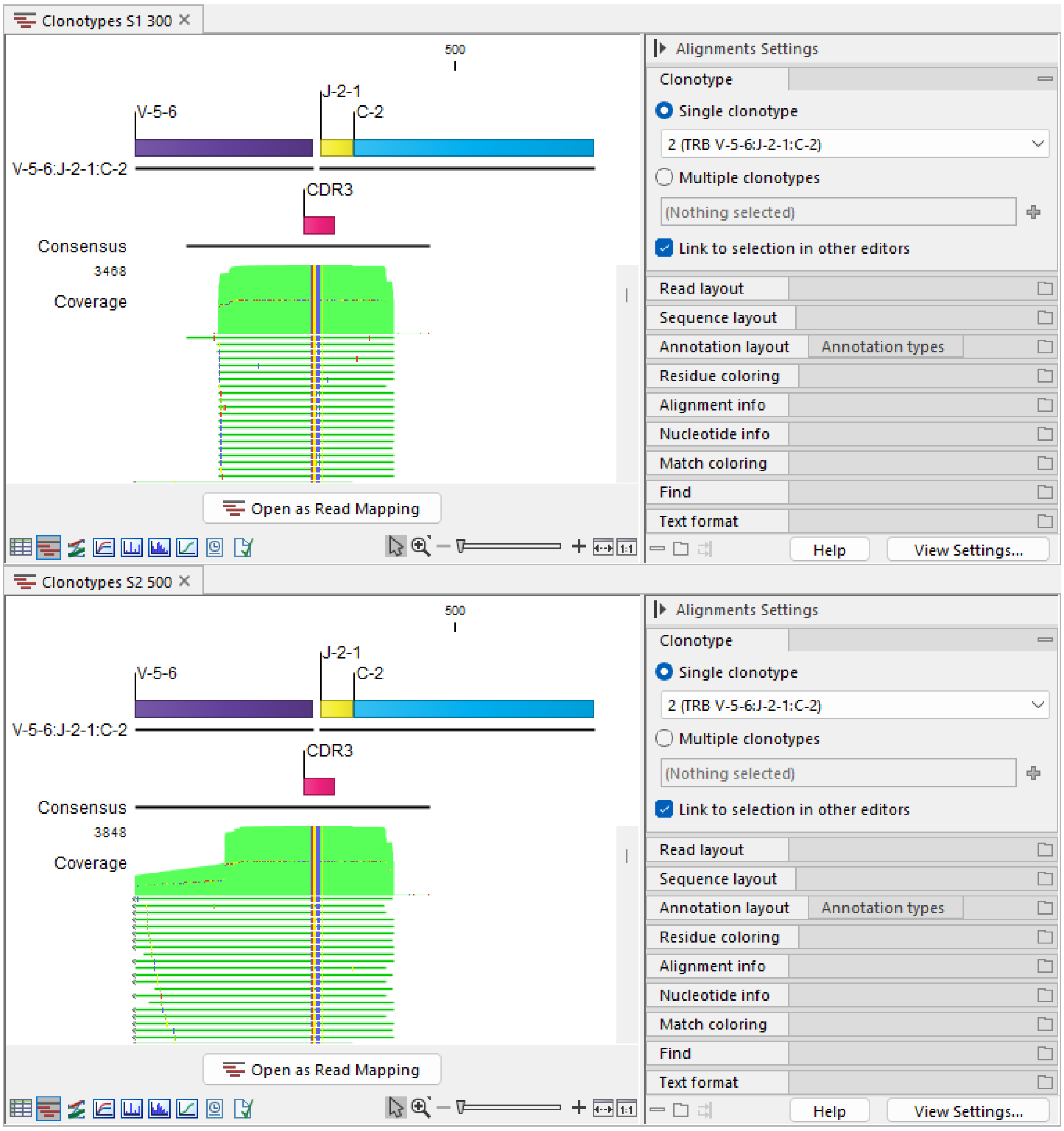


Figure 10: *Cumulative frequencies for the TRA chain from "Filter Immune Repertoire Report".*

Review the identified clonotypes

The primary view of the Clonotypes element is a table listing all unique identified clonotypes. This table can be filtered to focus on specific chains or productive clonotypes, depending on the analysis needs.

All report plots are also available as separate views within the **Clonotypes** element. In addition, the alignment view shows, for each clonotype, the reads assigned to it mapped against the corresponding gene segments, allowing assessment of V and C segment coverage (figure 11), as well as visualization of the random nucleotide insertions present at the junctions between segments in the CDR3 sequence (figure 12). This view (figure 11) clearly shows that the 500-cycle run produces longer reads, resulting in coverage of a larger portion of the V segment.



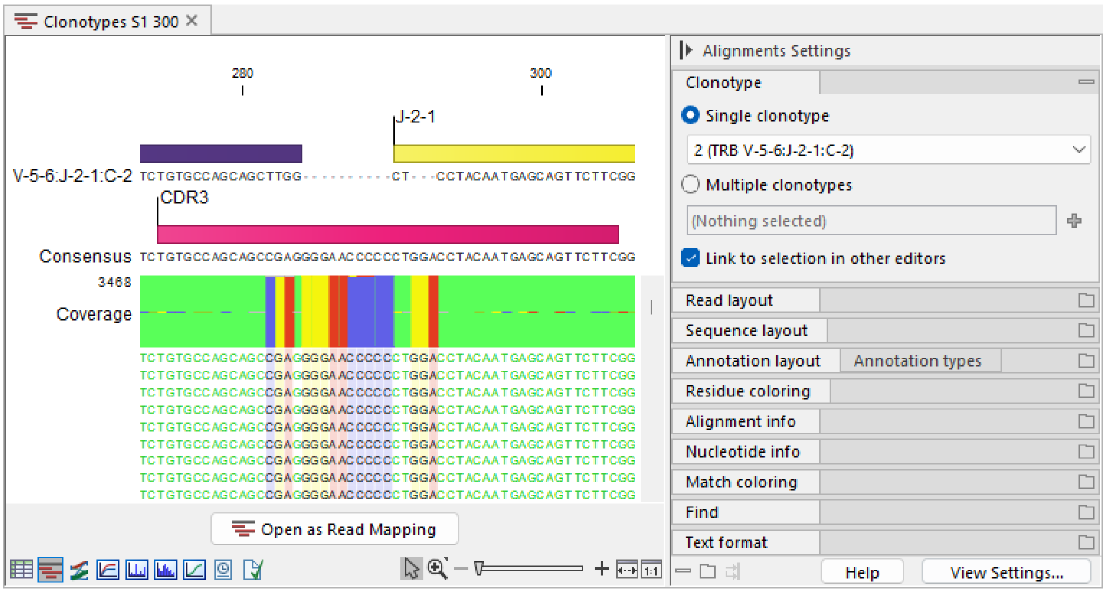
Clonotypes S1 300 – Clonotype:

● Single clonotype
2 (TRB V-5-6:J-2-1:C-2)
○ Multiple clonotypes
(Nothing selected)
☑ Link to selection in other editors

Clonotypes S2 500 – Clonotype:

● Single clonotype
2 (TRB V-5-6:J-2-1:C-2)
○ Multiple clonotypes
(Nothing selected)
☑ Link to selection in other editors

Figure 11: *Alignment view of the most frequent TRB clonotype, shown for both 300- and 500-cycle runs.*



● Single clonotype

2 (TRB V-5-6:J-2-1:C-2)

○ Multiple clonotypes

(Nothing selected)

☑ Link to selection in other editors

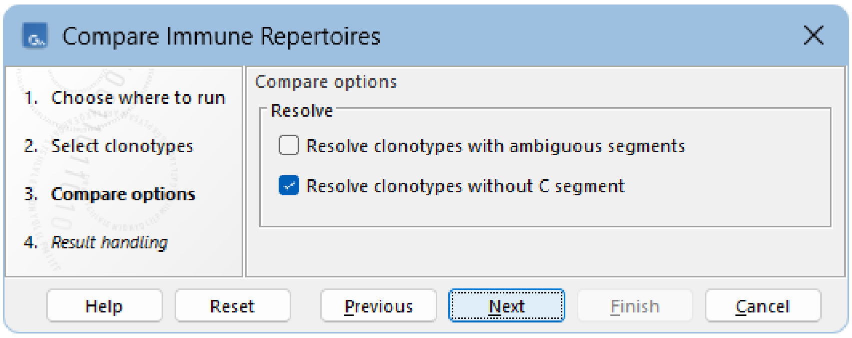
Figure 12: *Alignment view of the most frequent TRB clonotype for the 300-cycle run, zoomed in on the CDR3 sequence to better visualize the random nucleotide insertions at the junctions between gene segments.*

Compare immune repertoires

Once clonotypes have been detected in multiple samples, their immune repertoires can be further compared, both in terms of their statistics and properties as well as their composition.

First, we will create a Clonotype Sample Comparison.

1. Launch Compare Immune Repertoires using Quick Launch ( ).
2. In the next step, "Select clonotypes", select **Clonotypes S1 300** and **Clonotypes S1 500**.
3. In the next step, "Compare options", uncheck **Resolve clonotypes with ambiguous segments** (figure 13).



Compare options

Resolve

☐ Resolve clonotypes with ambiguous segments

☑ Resolve clonotypes without C segment

Figure 13: *Immune repertoires are compared without resolving ambiguous segments.*

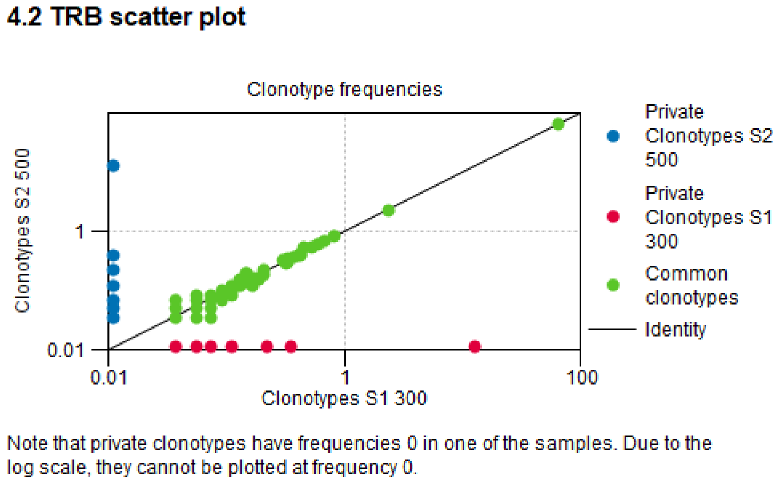
4. In the next step, "Result handling", select **Save**.
5. In the last step, choose to save the results in the "Results" folder and click on **Finish**.

The tool produces two outputs: a report and a **Clonotype Sample Comparison**.

The report includes sections analogous to those in the single-sample reports, containing data from both samples to enable direct comparison.

The **Resolved clonotypes** section summarizes the number of clonotypes that have been resolved. Resolving clonotypes across samples is conceptually similar to merging clonotypes within a single sample, where information from one sample can be used to assign missing C segments or disambiguate V, D, J, and C segments in other samples. In this tutorial, ambiguous segments have not been resolved, to illustrate the impact of the shorter R2 read in the 300-cycle run.

The **Scatter plots** section contains a scatter plot for each chain, comparing clonotype frequencies between the two samples, allowing assessment of the repertoires similarity (figure 14). Each run contains one private clonotype with frequency above 1%. These cloonotypes can be further inspected using the **Clonotype Sample Comparison**.



4.2 TRB scatter plot

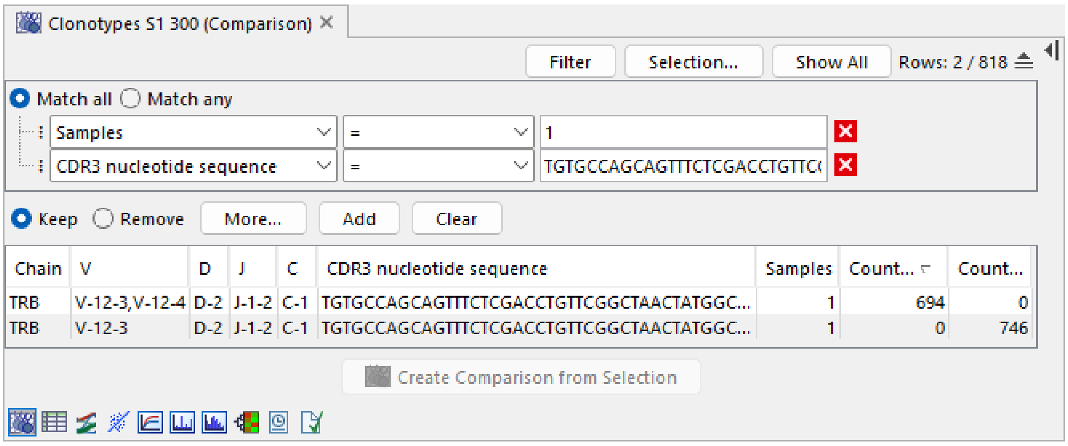
Note that private clonotypes have frequencies 0 in one of the samples. Due to the log scale, they cannot be plotted at frequency 0.

Figure 14: *Scatter plot for the TRB chain. Blue and red dots represent clonotypes found in only one run, while green dots indicate clonotypes present in both. As expected for two sequencing runs of the same sample, the repertoires are highly similar, with many shared clonotypes at comparable frequencies.*

The primary view of the Clonotype Sample Comparison element is a table listing all unique clonotypes across all samples, along with their counts and frequencies in each sample. All report plots are also available as separate views within this element. The table can be filtered (figure 15) to identify private clonotypes with high frequency by:

1. Filtering using *Samples* = *1* to show clonotypes present in only one sample.
2. Sorting by count or frequency in one of the samples to highlight high-frequency private clonotypes.
3. Filtering using the CDR3 nucleotide sequence of the highest-frequency clonotype. We recommend the right-click menu for applying this filter.

The table reveals that these high-frequency private clonotypes are in fact the same clonotype, but the shorter R2 read in the 300-cycle run results in an ambiguous V segment assignment. Checking **Resolve clonotypes with ambiguous segments** resolves these entries to a single clonotype.



Rows: 2 / 818

● Match all ○ Match any

Samples = 1

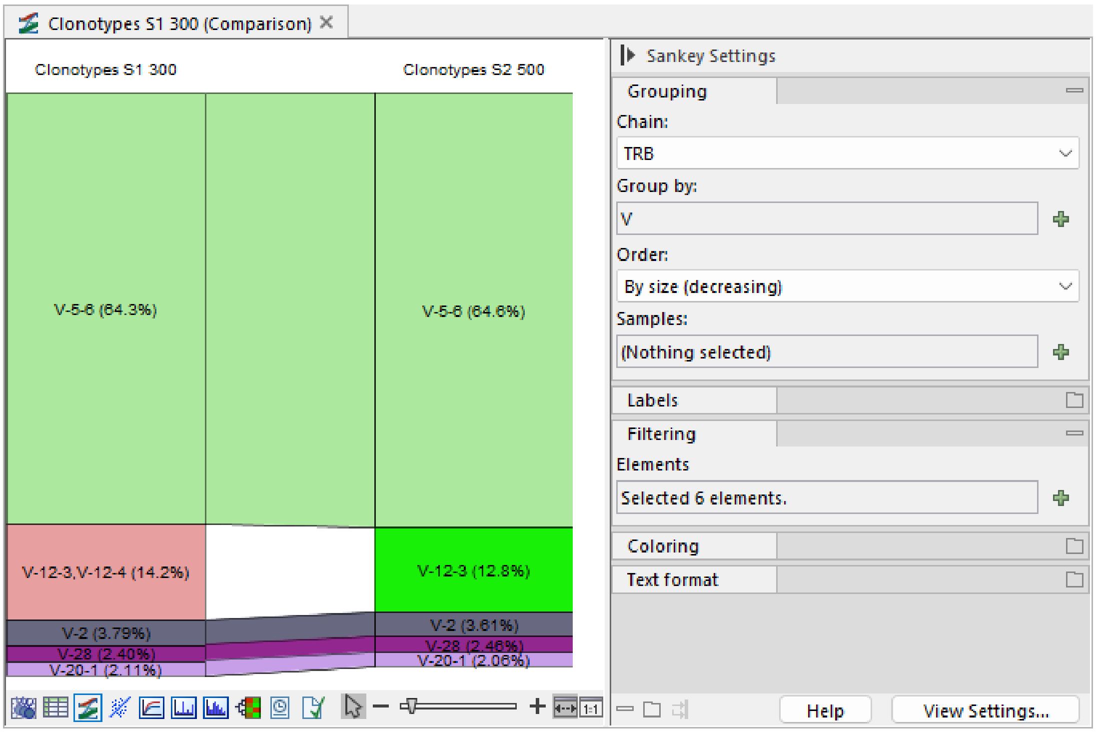
CDR3 nucleotide sequence = TGTGCCAGCAGTTTCTCGACCTGTTC(

● Keep ○ Remove

Chain	V	D	J	C	CDR3 nucleotide sequence	Samples	Count...	Count...
TRB	V-12-3,V-12-4	D-2	J-1-2	C-1	TGTGCCAGCAGTTTCTCGACCTGTTCGGCTAACTATGGC...	1	694	0
TRB	V-12-3	D-2	J-1-2	C-1	TGTGCCAGCAGTTTCTCGACCTGTTCGGCTAACTATGGC...	1	0	746

Figure 15: *Table view for the Clonotype Sample Comparison. The table is filtered to show private clonotypes that share a specific CDR3 nucleotide sequence. Note that not all table columns are shown.*

The Sankey plot view illustrates the same pattern: clonotypes with ambiguous V segment assignments appear as private to the 300-cycle run, while the corresponding clonotypes with unambiguous V segments appear as private to the 500-cycle run.



Chain: TRB

Group by: V

Order: By size (decreasing)

Samples: (Nothing selected)

Elements: Selected 6 elements.

Figure 16: *Sankey plot view for the Clonotype Sample Comparison. The plot shows the V segments for the TRB chain and is filtered to show only the five most frequent V segments in each run.*

When comparing more than two samples, the Jaccard distance heat map can be useful for visualizing the similarity across samples and how they cluster together.