Tutorial

Calculate TMB Score and MSI Status

March 5, 2026

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

Sample to Insight

Tutorial

Calculate TMB Score and MSI Status

The purpose of this tutorial is to demonstrate how *CLC Genomics Workbench* and the *Biomedical Genomics Analysis* plugin can be used to calculate:

- A Tumor Mutation Burden (TMB) score. High TMB scores have been shown to be correlated with improved patient response rates to immune check-point inhibitors [A. Stenzinger et al., 2019].
- A MicroSatellite Instability (MSI) status. Many patients with MSI-high tumors have shown favorable responses to immunotherapy. Furthermore, the U.S. Food and Drug Administration (FDA) granted accelerated approval for pembrolizumab in both pediatric and adult patients with MSI-high tumors, marking the first treatment approved for any solid tumor based on this biomarker, regardless of tumor origin [Chang et al., 2018].

Although both TMB and MSI serve as genomic biomarkers linked to immunotherapy response, they reflect distinct biological processes and are not inherently correlated. MSI results from defects in the DNA mismatch repair system, whereas TMB can be elevated due to various factors, such as environmental mutagens or alterations in DNA repair genes. A large cohort study of over 40,000 cancer samples [Frampton et al., 2016] found that while most MSI-high tumors also exhibited high TMB, a substantial proportion of high-TMB tumors were microsatellite stable (MSS) in specific cancer types, such as lung, skin, and urinary tract. This highlights the importance of assessing both TMB and MSI independently.

In this tutorial, we focus on the following:

- Import data.
- Calculate a TMB score and MSI status using a template workflow.
- Interpret the results.

Data used in this tutorial

This tutorial uses data produced with a QIAseq Targeted DNA Panel.

To complete the tutorial in a reasonable amount of time, only a subset of reads from a single sample that map to chromosome 22 are used here.

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.

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).

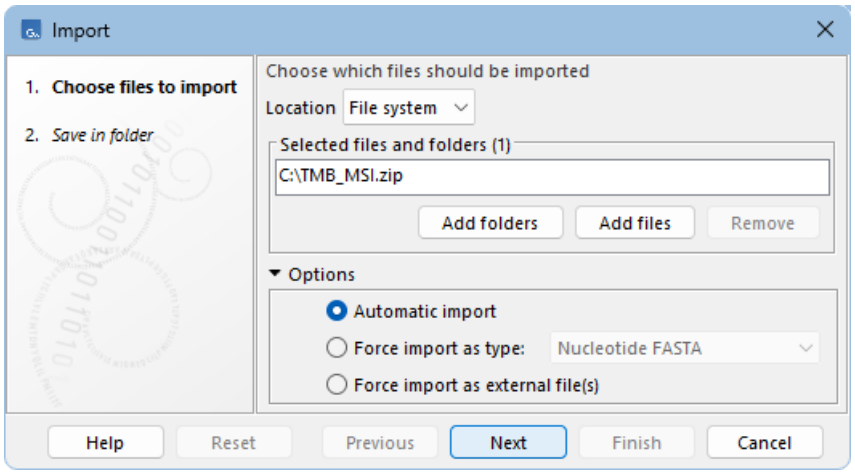

Figure 1: *Standard Import configured to import the tutorial data.*

   (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, a "TMB_MSI" folder containing the S11 sample reads is visible in the Navigation Area.

Calculate TMB score and MSI status

We will now use the Analyze QIAseq DNA Somatic (Illumina) template workflow to analyze the tutorial data and calculate a TMB score and an MSI status. This workflow has been designed for data generated using a QIAseq Targeted DNA 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 content of the workflow, locate it in the Toolbox:

> **Workflows** | **Template Workflows** | **Biomedical Workflows** ( ) | **QIAseq Sample Analysis** ( ) | **QIAseq DNA workflows** ( ) | **Analyze QIAseq DNA Somatic (Illumina)** ( )

Then 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 **Specify workflow path** step, set "Calculate TMB" and "Detect MSI" to "Yes" (figure 2).

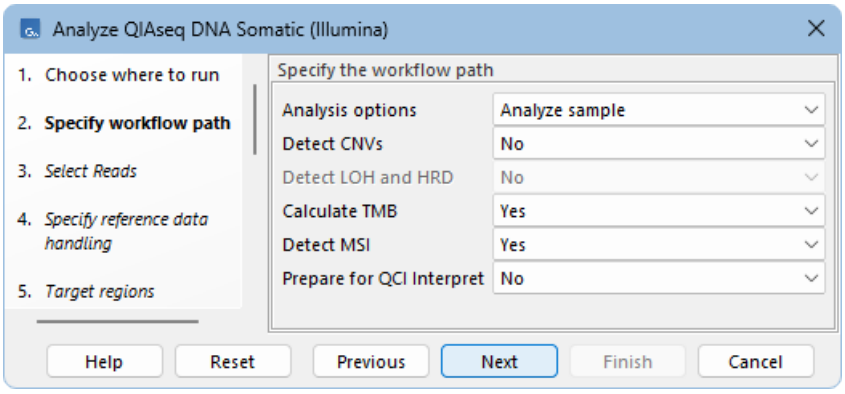

Figure 2: *Configure workflow to calculate TMB score and detect MSI.*

3. In the **Select Reads** step, specify the data to be analyzed by selecting the imported reads (figure 3).

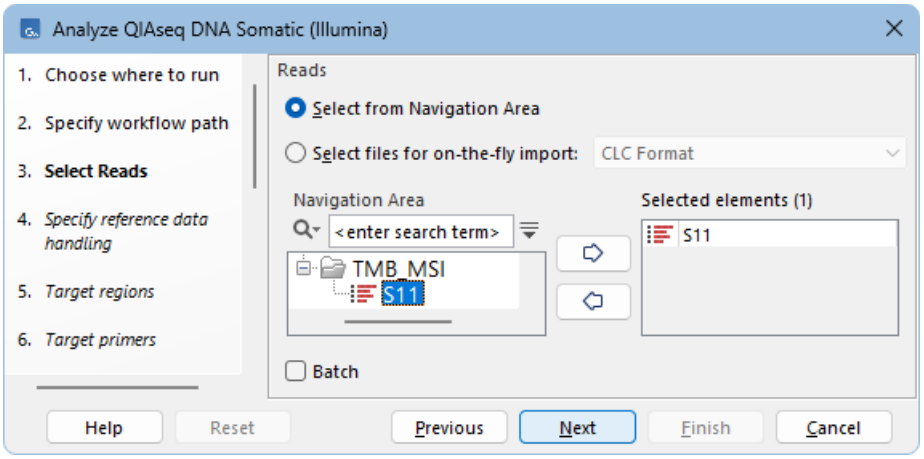

Figure 3: *Select the reads to use as input.*

4. In the **Specify reference data handling** step, select the "TMB and MSI" reference data set under "QIAGEN Tutorial" (figure 4). Click **Download to Workbench** if the data set has not already been downloaded.
5. In the next steps, keep the default settings.

   The thresholds for calculating the TMB status from the TMB score can be adjusted as needed in the Calculate TMB Score step.
6. In the last step, make a new subfolder in "TMB_MSI" called "Results" and choose to save the workflow results there.

   Click on **Finish**.

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

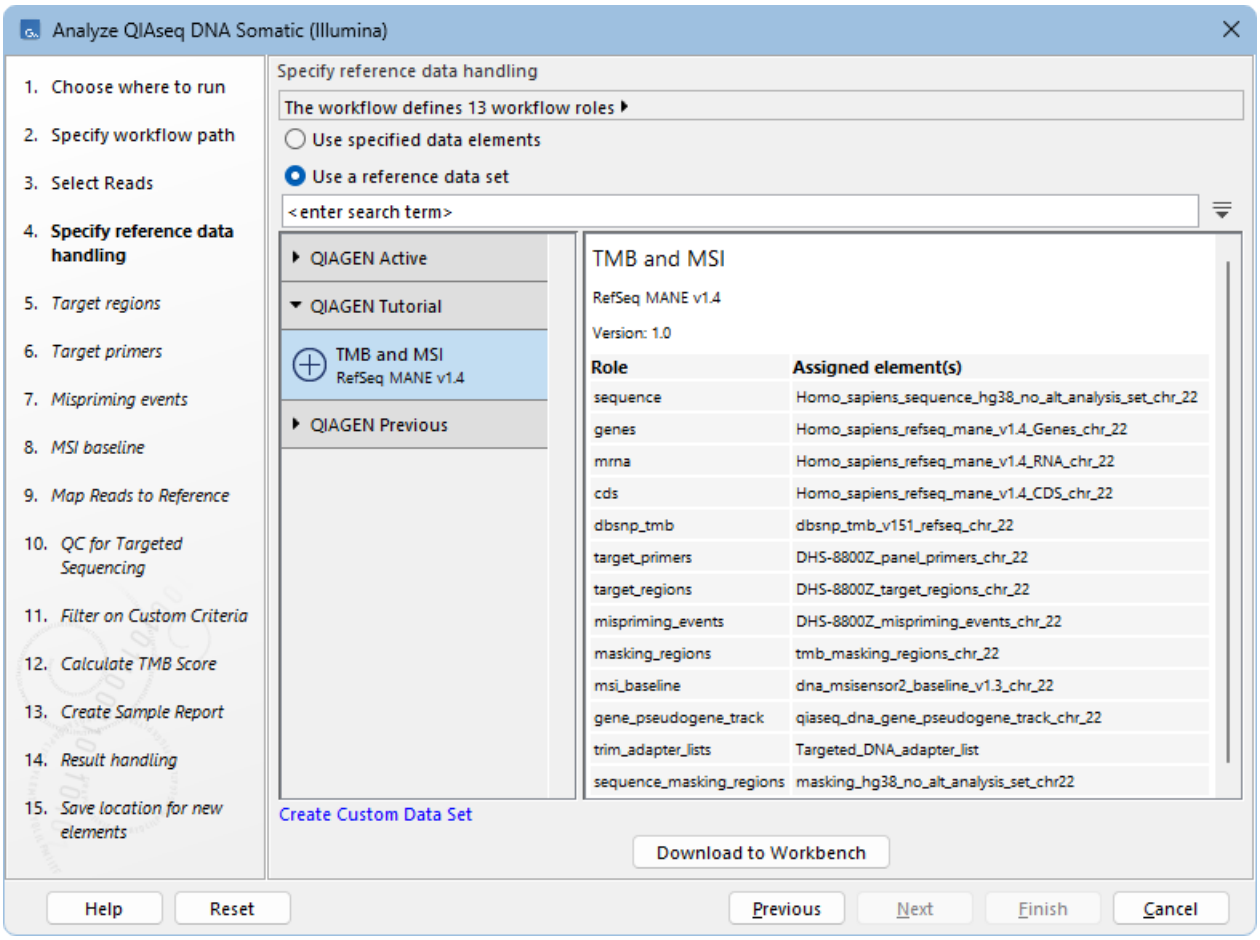

Figure 4: *Select the tutorial reference data set.*

Interpret the results

Results from the workflow are placed in the selected location (figure 5).

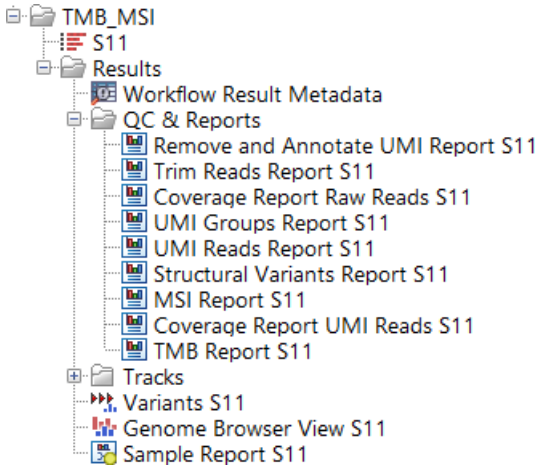

Figure 5: *The "Results" folder in the Navigation Area.*

The folder contains, among other things:

- A **QC & Reports** subfolder containing all reports produced by the workflow.
- A **Tracks** subfolder containing, among other things, the variants used for computing the TMB score, and the MSI loci.
- A Sample Report summarizing information from all the reports located in the QC & Reports folder.

  The Sample Report is a good place to start to get a comprehensive **overview of the QC metrics** for the analyzed data, as well as the calculated TMB score and MSI status. See details in the section below.

The Sample Report

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

Open the **Sample Report S11**, found in the "Results" 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 reads mapped in target region	Raw read coverage	90.02	≥ 50.00
Percentage of target region positions with coverage ≥ threshold	UMI read coverage	88.11	≥ 90.00

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

Here, the "Quality control" section shows that the "Percentage of target region positions with coverage $\geq$ threshold" did not pass the predefined threshold and is therefore colored yellow (figure 6). This is expected because the data was downsampled to chromosome 22. Therefore, this does not indicate a problem with the analysis.

TMB score

The Sample Report contains a "Calculate TMB score" section, summarizing the main TMB results, including the score and status (figure 7).

The TMB score is calculated as the number of non-synonymous somatic variants per Mb assessed regions. The workflow provides as input to the tool the target regions with at least 100x coverage. As the tutorial data is restricted to chromosome 22, the length of the target regions is too short, and therefore the TMB calculations are not reliable. This can be seen in more details in the **QC & Reports/TMB Report S11** report, where the status is highlighted in red.

MSI status

The Sample Report contains a "Detect MSI status" section, summarizing the main MSI results, including the number of unstable loci and status (figure 8).

11 Calculate TMB score

11.1 Summary

Report	TMB Report S11
TMB status	High
Length of target regions (bp)	20,867
Variants inside target regions and after quality filters	5
Germline variants	4
Somatic variants	1
Non-coding somatic variants	0
Synonymous somatic variants	0
Non-synonymous somatic variants	1
Tumor mutational burden (mutations/Mb)	47.92

Figure 7: *The summary TMB results in the Sample Report.*

12 Detect MSI status

Report	MSI Report S11
Loci (#)	2
Unstable loci (#)	0
Stable loci (#)	2
Not testable loci (#)	0
Unstable loci (%)	0
MSI status	MSS
Clinical term	MS-stable

Figure 8: *The summary MSI results in the Sample Report.*

The sample is classified as MS-stable because all loci are evaluated as stable, with each locus showing a length distribution similar to the baseline. The distributions can be investigated in more details in the **QC & Reports/MSI Report S11** report (figure 9).

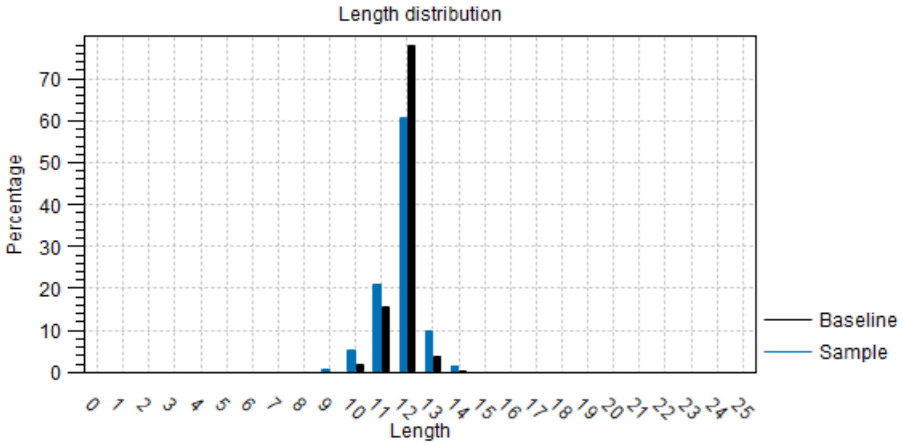

Figure 9: *A locus shows a similar distribution in the sample (blue) and baseline (black).*

As the tutorial data is restricted to chromosome 22, only two loci are used for the calculations. For your own data, we do not recommend assessing MSI status using fewer than five loci.

Bibliography

[A. Stenzinger et al., 2019] A. Stenzinger, J. A., Maas, J., Stewart, M., Merino, D., Wempe, M., and Dietel, M. (2019). Tumor mutational burden standardization initiatives:recommendations for consistent tumor mutational burdenassessment in clinical samples to guide immunotherapytreatment decisions. *Genes C*, pages 1–11.

[Chang et al., 2018] Chang, L., Chang, M., Chang, H., and Chang, F. (2018). Microsatellite instability: A predictive biomarker for cancer immunotherapy. *Applied I*, 26(2):e15–e21.

[Frampton et al., 2016] Frampton, G., Fabrizio, D., Chalmers, Z., Sun, J., Miller, V., and Stephens, P. (2016). Assessment and comparison of tumor mutational burden and microsatellite instability status in> 40,000 cancer genomes. *Annals of Oncology*, 27:vi15.