Tutorial

Analyze CGP Data with LightSpeed

January 15th, 2025

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

Sample to Insight

Tutorial

Analyze CGP Data with LightSpeed

Introduction

The purpose of this tutorial is to demonstrate how *CLC Genomics Workbench* and the *LightSpeed* module can be used to analyze comprehensive genomic profiling (CGP) DNA panel data.

This is illustrated using Twist CGP data and a corresponding template workflow. Most steps described here are directly transferable to QIAseq CGP data as well, for which the following template workflows are available in the *LightSpeed* module:

- QIAseq xHYB CGP Fastq to Somatic Variants
- QIAseq xHYB CGP cfDNA Fastq to Somatic Variants

In this tutorial, we focus on the following:

- Processing CGP data using a template workflow.
- Assessing quality control metrics.
- Interpreting results.

Data used in this tutorial

This tutorial uses data produced from Horizon's Structural Multiplex FFPE Reference Standard (HD789) using Twist Oncology DNA CGP Panel.

The sample contains somatic variants, DNA fusions and copy number variations (CNVs). It has been sequenced to approx. 600x mean target coverage to demonstrate detection of all known events.

Prerequisites

For this tutorial, you must be working with *CLC Genomics Workbench* 26.0.1 and *LightSpeed* module 26.0 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.
- 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.

Process the data

Start by downloading the tutorial data. Launch the *CLC Genomics Workbench* with the *LightSpeed* module installed.

We will use the Twist CGP Fastq to Somatic Variants template workflow to process the tutorial data.

This workflow has been designed for data generated using the Twist Oncology DNA CGP Panel without unique molecular indexes (UMIs). In this tutorial, we will also cover how to adjust the workflow to accomodate the following:

- Use of Twist UMIs.
- Use of a custom panel.
- Use of own control samples for creation of CNV and MSI baselines.

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 application.

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

> **Template Workflows** | **LightSpeed Workflows** ( ) | **Twist Workflows** ( ) | **Twist CGP Fastq to Somatic Variants** ( )

Right-click on its name and choose **Open Copy of Workflow** (figure 1).

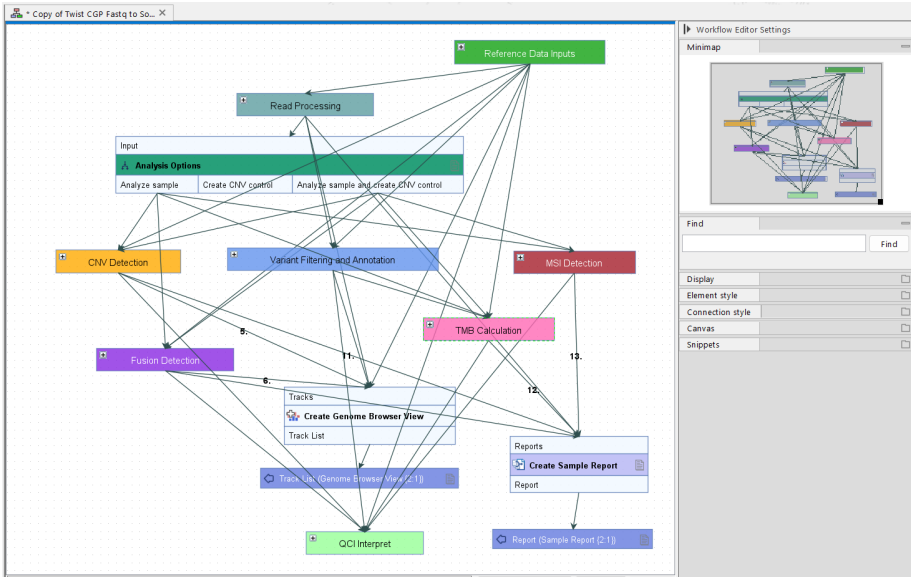


Figure 1: *Overview of the template workflow.*

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 step, choose where to run the workflow.

3. In the "Specify workflow path" step, select the intended type of analysis and the desired features. Choose the default option (figure 2).
4. In the "Specify reference data handling" step, choose the appropriate reference data. The *Twist CGP hg38* reference data set will be preselected and is appropriate for the analysis (figure 3). Click the **Download to Workbench** button unless the reference data set has already been downloaded.
5. In the "LightSpeed Fastq to Somatic Variants" step, click the **Browse** button and locate the fastq files downloaded for this tutorial (figure 4).
6. For the purpose of this tutorial, continue with default settings in the remaining wizard steps.
7. In the final step, choose where to save the workflow results.

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

Note that on first use of a reference sequence, the runtime is increased because a reference index is generated. This is cached and subsequent runs will therefore be faster.

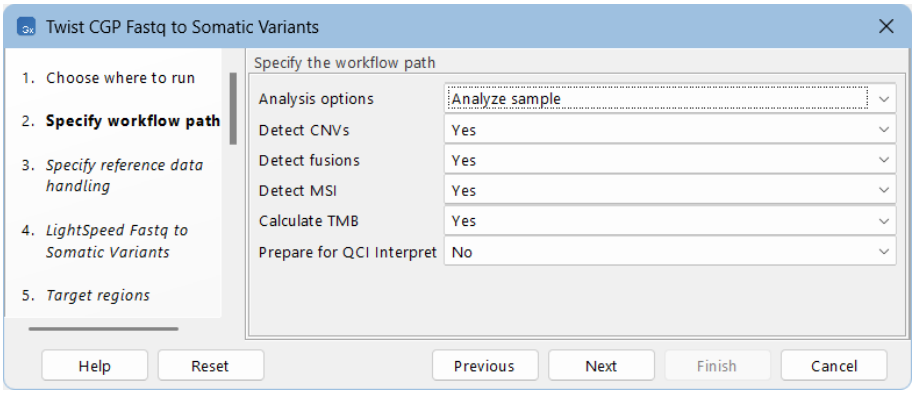


Figure 2: *Specify the analyses to be carried out.*

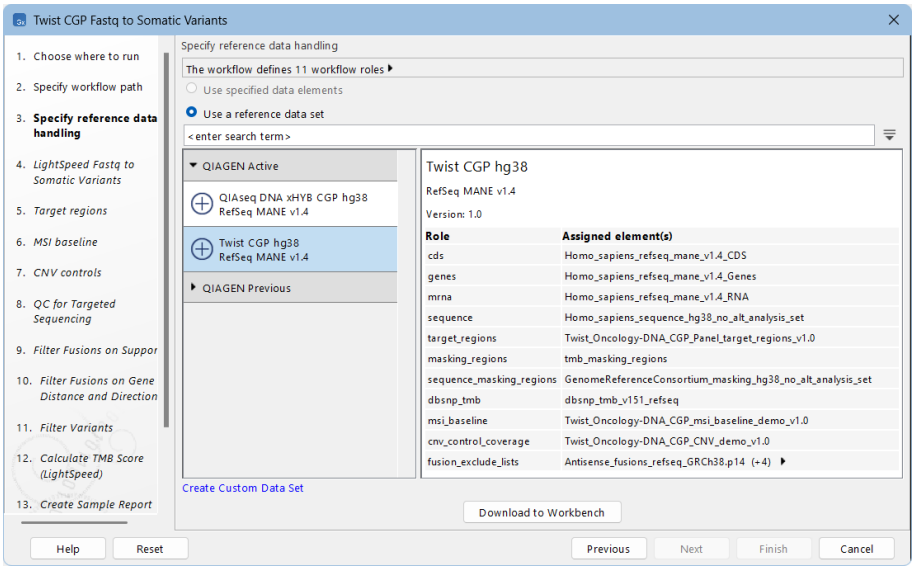


Figure 3: *The Twist CGP hg38 reference data set is selected.*

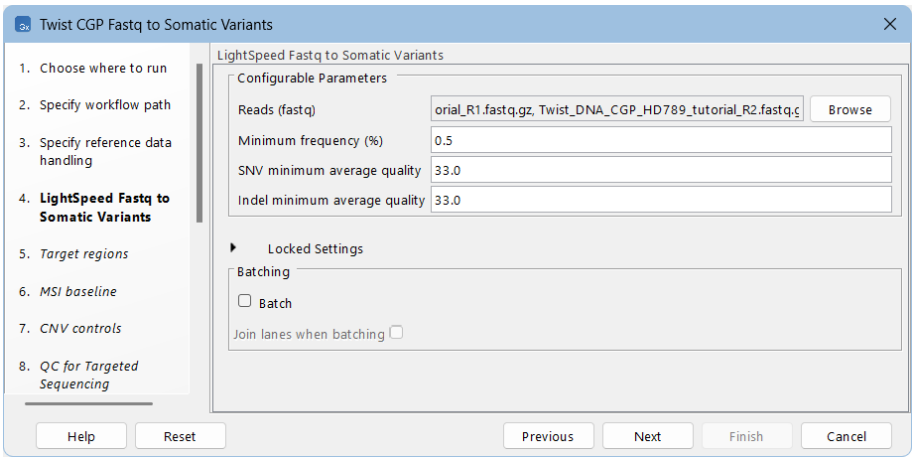


Figure 4: *Locate the tutorial data fastq files for analysis.*

Inspect the workflow outputs

The workflow will produce all relevant outputs to assess QC metrics and interpret results (figure 5):

- A **QC & Reports** folder containing all reports produced by the workflow including TMB and MSI analysis.
- A **Tracks** folder containing various tracks for in-depth assessment of the results including a read mapping.
- A **Genome Browser View** ( ) with relevant results and annotations included.
- A **Sample Report** ( ) summarizing information from all the reports located in the QC & Reports folder.
- Three tracks with the primary track results of the workflow, **Variants**, **Gene-level CNV Track**, and **Fusion Genes**.

Assess QC metrics

Open the **Sample Report**. The Quality control section contains different summary items that can be used to assess the quality of the data (figure 6). These summary items can be configured in the "Create Sample Report" wizard step when launching the workflow. Notice that "Percentage of target region positions with coverage $\geq$ threshold" is marked in yellow to highlight that the metric did not meet the specified value which was more than 90% of target region positions to have at least 400x coverage.

Go to section 2.1 of the Sample Report to inspect summary items such as "Input read pairs", "Median insert size" and "Read pairs remaining for variant detection". Additional details are available in the sections below. To exemplify, go to section "2.13 Deduplication" and verify that the "Duplicate read pairs (% of mapped)" is at approx. 7% (figure 7).

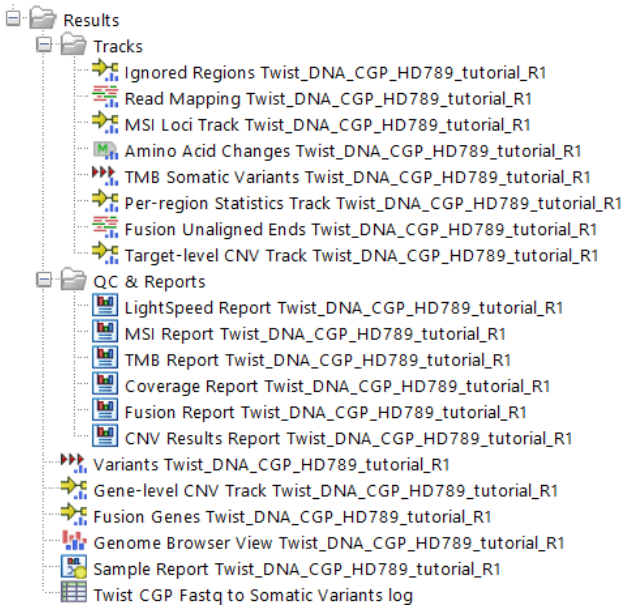


Figure 5: *The outputs generated by the workflow.*

1.2 Quality control

Summary item	Report type	Value	Threshold
Percentage of target region positions with coverage ≥ threshold	QC for targeted sequencing	75.71	≥ 90.00
Percentage reads mapped in target region	QC for targeted sequencing	78.11	≥ 50.00

Figure 6: *The quality control section of the Sample Report. Note that the summary items are displayed with traffic light coloring to provide an easy overview of selected QC metrics.*

2.13 Deduplication

Report	LightSpeed Report Twist_DNA_CGP_HD789_tutorial_R1
Duplicate read pairs (% of mapped)	6.96
Read pairs after deduplication (% of mapped)	93.04
Broken read pairs after deduplication (% of mapped)	0.39
Non-specific read pairs after deduplication (% of mapped)	2.60
Specific read pairs after deduplication (% of mapped)	97.01

Figure 7: *The deduplication section of the Sample Report presenting deduplication metrics.*

Interpret the results

Here, we will review the primary results of the workflow.

Somatic variants

We will now verify we were able to call the expected somatic variants.

- Open the **Genome Browser View**.
- In the side panel of the Genome Browser View, in "Track List Settings", go to chromosome 19 and type "3118944" in the "Location" field. This will bring the genome browser view to an expected variant, GNA11 Q209L, and highlight it in the variant table below (figure 8). The variant annotation column "Amino acid change" provides information about the observed amino acid change.

- Continue verifying that the expected variants are called:
  - Chr 3:179218303 (PIK3CA E545K)
  - Chr 7:55174771 (EGFR E746-A750del)
  - Chr 7:55181317 (EGFR V769_D770insASV)
  - Chr 14:104780214 (AKT1 E17K)

Figure 8: *Split view showing the genome browser at the top and a table view of the identified variants at the bottom. The GNA11 Q209L variant is highlighted.*

CNVs

We will now verify we were able to call the expected gene CNVs.

Locate the **Gene-level CNV track** in the Genome Browser View, hover the indicator over it and click "Open table view". This will open a table view of the identified gene-level CNVs (figure 9). Two CNVs are expected:

- MET amplification (4.5 copies)
- MYCN amplification (8.5 copies)

Verify in the table view that these are called. We can further verify the called copy number by looking at the "Fold-change" column and multiplying by 2. MET was called at 5.24 copies and MYCN was called at 9.88 copies.

A graphical representation of the identified CNVs is available in the **CNV Results Report**.

DNA fusions

We will now verify we were able to call the expected DNA fusions.

Locate the **Fusion Genes** track in the Genome Browser View, hover the indicator over it and click "Open table view". This will open a table view of the identified DNA fusions after post filtering (figure 10). The final DNA fusion filtering step in the template workflow does not remove DNA

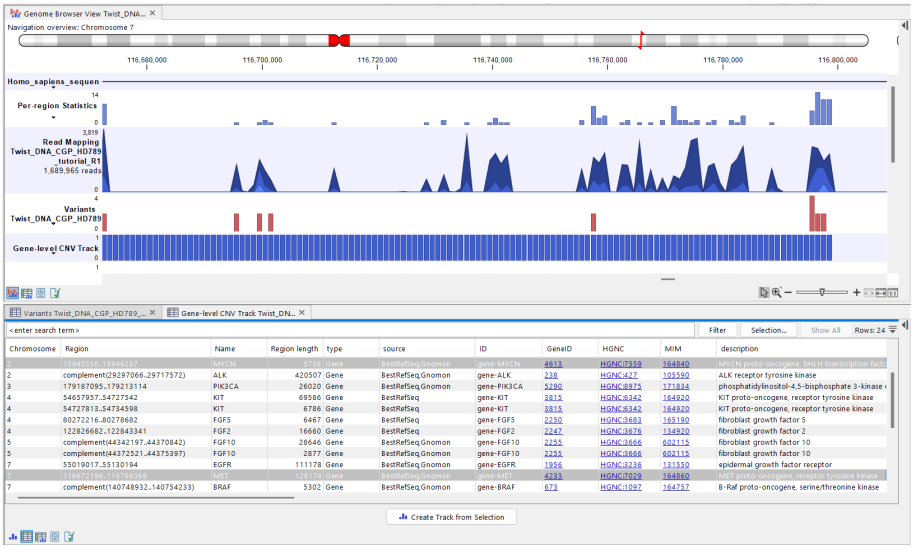


Figure 9: *Split view showing the genome browser at the top and a table view of the identified CNVs at the bottom. The MET amplification is highlighted in the genome browser view and both expected CNVs are highlighted in the table view.*

fusions but instead annotates them based on presence in Ingenuity Pathway Analysis (IPA). This is indicated by the columns "IPA gene view" and "Filter". Two DNA fusions were identified in the data that passed the final IPA filter:

- SLC34A2-ROS1
- CCDC6-RET1

These are the two expected DNA fusions.

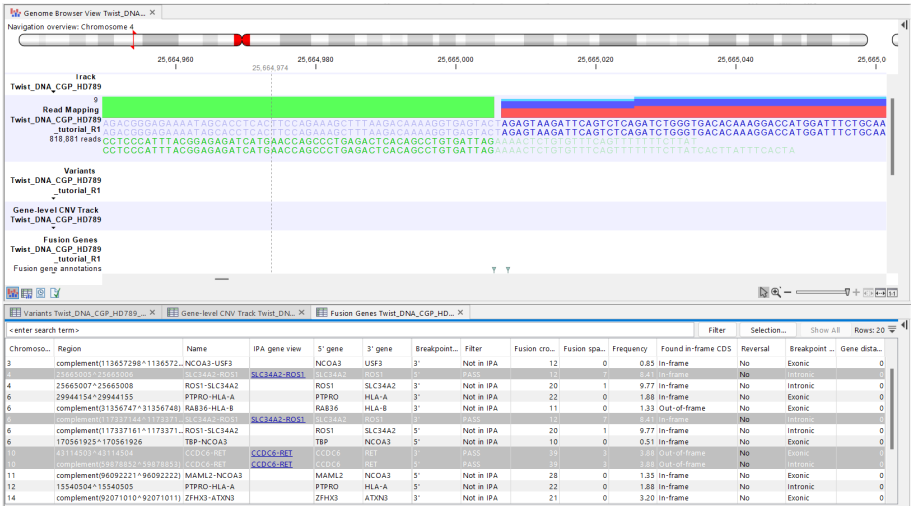
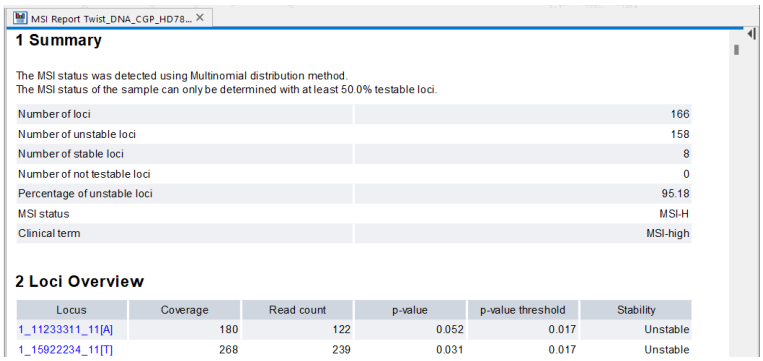


Figure 10: *Split view showing the genome browser at the top and a table view of the identified DNA fusions at the bottom. One breakpoint of the SLC34A2-ROS1 fusion is highlighted in the genome browser view and both expected DNA fusions are highlighted in the table view.*

MSI and TMB

Although there are no expected MSI and TMB scores for the HD789 sample, we will review these results for completeness.

Go to the QC & Reports folder and open the **MSI Report** (figure 11). In the Summary section, we are presented with information about the number of microsatellite loci that were investigated, how many were found to be stable/unstable/non-testable, the percentage of unstable loci, i.e., the MSI score, and the resulting MSI status.


MSI Report Twist_DNA_CGP_HD78...

1 Summary

The MSI status was detected using Multinomial distribution method.
The MSI status of the sample can only be determined with at least 50.0% testable loci.

Number of loci	166
Number of unstable loci	158
Number of stable loci	8
Number of not testable loci	0
Percentage of unstable loci	95.18
MSI status	MSI-H
Clinical term	MSI-high

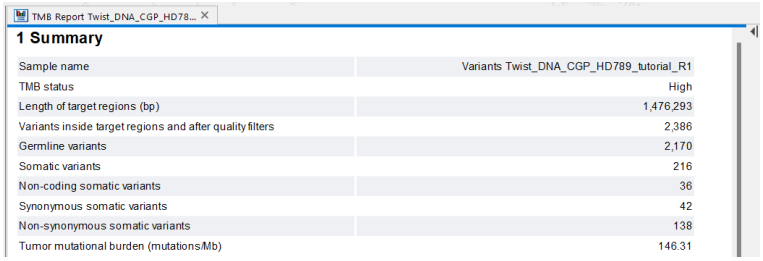
2 Loci Overview

Locus	Coverage	Read count	p-value	p-value threshold	Stability
1_11233311_11[A]	180	122	0.052	0.017	Unstable
1_15922234_11[T]	268	239	0.031	0.017	Unstable

Figure 11: *The MSI report showing the MSI status of the sample and metrics for the analyzed loci.*

Additional metrics and plots illustrating the length distributions per loci are available in the sections below.

Open the **TMB Report** (figure 12). In the Summary section, we are presented with information about the number of variants for TMB analysis, the size of the targets used in the analysis, the Tumor mutational burden, and the resulting TMB status.


TMB Report Twist_DNA_CGP_HD78...

1 Summary

Sample name	Variants Twist_DNA_CGP_HD789_tutorial_R1
TMB status	High
Length of target regions (bp)	1,476,293
Variants inside target regions and after quality filters	2,386
Germline variants	2,170
Somatic variants	216
Non-coding somatic variants	36
Synonymous somatic variants	42
Non-synonymous somatic variants	138
Tumor mutational burden (mutations/Mb)	146.31

Figure 12: *The TMB report showing the TMB score and status of the sample and metrics for the analyzed somatic variants.*

Additional information regarding variant filtering and observed variant frequencies is available in the sections below.

Modifying the workflow

In this section, we will describe how to modify the workflow to accomodate changes in data generation or analysis setup. Specifically, we will cover the following scenarios:

- Use of Twist UMIs.
- Use of a custom panel.
- Use of own control samples for creation of CNV and MSI baselines.

Use of Twist UMIs

The Twist CGP Fastq to Somatic Variants template workflow can with a few steps be configured to support CGP data generated with Twist UMIs or a custom UMI read structure. To configure the workflow:

1. Locate the workflow in the Toolbox, right-click on the workflow and click **Open copy of workflow**. This will open a copy of the workflow presenting a detailed view of all the included tools and components.
2. Type "LightSpeed fastq" In the "Find" field in the side panel. This will highlight the **LightSpeed Fastq to Somatic Variants** tool. Double-click it.
3. Go to the "Reads" wizard step and click on the drop-down menu for "UMI preset".
   - For Twist duplex UMI, select **Twist UMI** (figure 13).
   - For a custom UMI read structure, select **Custom** and specify UMI and common sequence lengths below.
4. Click **Finish**.
5. Save the workflow.

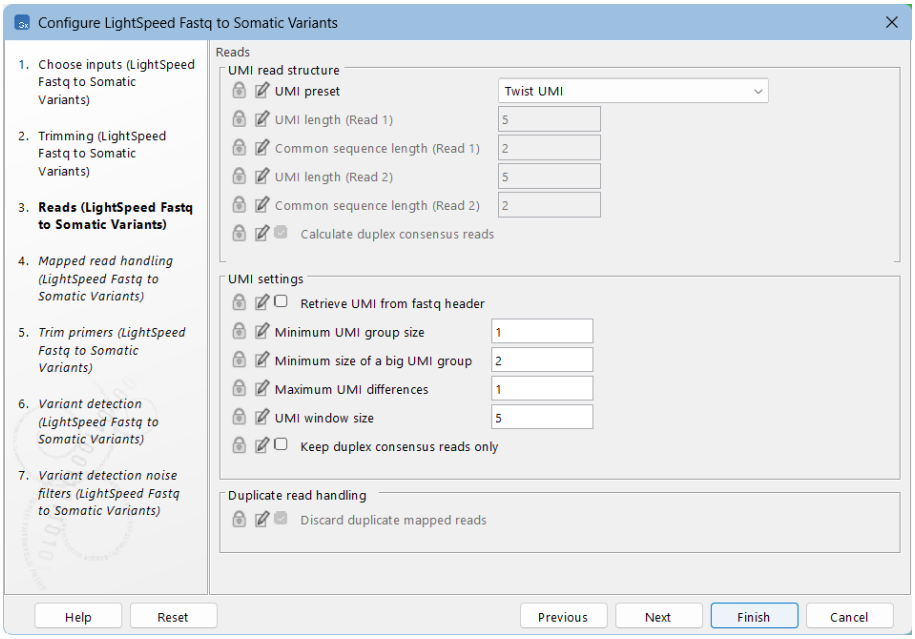


Figure 13: *Twist UMI configured in the LightSpeed Fastq to Somatic Variants tool for proper handling of Twist duplex UMIs.*

Use of a custom panel

Twist Oncology DNA CGP panel target regions are available in the *CLC Genomics Workbench* reference data. If custom target regions have been added or a custom panel has been used, the Twist CGP Fastq to Somatic Variants template workflow can be configured to support that. To enable processing of data created with a custom panel:

1. Import the target region bed file.
    - Click **Import** and select **Tracks...**
    - In the first step, choose where to run the import.
    - Click the drop-down menu for "Type of files to import" and select **BED**.
    - Click **Add files** and browse to the target region bed file.
    - Click the **Browse** button next to the "Reference track" field.
    - Go to the "Reference Data" tab, go to QIAGEN Active | hg38 (Ensembl) | sequence and select **Homo_sapiens_sequence_hg38_no_alt_analysis_set**.
    - In the next wizard step, choose to save the results.
    - In the final wizard step, choose where to save the results.
2. Locate the workflow in the Toolbox, right-click on the workflow and click **Open copy of workflow**.
3. Locate the **Target regions** input in the workflow group "Reference Data Inputs". Double-click it and delete the text from the "Workflow role" field.
4. Click Browse next to the "Workflow input" field and select the imported custom target regions in the Navigation Area. Click OK.
5. Save the workflow.

Use of own control samples for creation of CNV and MSI baselines

CNV and MSI baselines are available for analysis of Twist Oncology DNA CGP panel data for demonstration purposes in the *CLC Genomics Workbench* reference data. We recommend processing your own control samples for use as CNV and MSI baselines to properly capture potential biases introduced during sample and/or library preparation.

To use your own control samples for CNV detection:

1. Launch the Twist CGP Fastq to Somatic Variants template workflow.
2. In the "Specify workflow path" wizard step select **Create CNV control** in the Analysis options dropdown menu.
3. In the "LightSpeed Fastq to Somatic Variants" wizard step, select the fastq files for the control samples. Tick **Batch** in the Batching options and tick **Join lanes when batching** if applicable.
4. Use default settings for all other wizard steps.
5. Once the analyses are complete, locate the workflow in the Toolbox, right-click on the workflow and click **Open copy of workflow**.
6. Locate the **CNV controls** input in the workflow group "CNV Detection". Double-click it and delete the text from the "Workflow role" field.
7. Click **Browse** next to the "Workflow input" field and select the coverage tables from the completed control sample analysis outputs in the Navigation Area. Click **OK**.

8. Save the workflow.

Note that CNV controls should be gender matched to the test sample at hand.

To use your own control samples for MSI analysis:

1. Skip step 2-5 if control samples have already been processed, e.g., for use as CNV control samples, as described above.
2. Launch the Twist CGP Fastq to Somatic Variants template workflow.
3. In the "Specify workflow path" wizard step select **Create CNV control** in the Analysis options drop-down menu.
4. In the "LightSpeed Fastq to Somatic Variants" wizard step, select the fastq files for the control samples. Tick **Batch** in the Batching options and tick **Join lanes when batching** if applicable.
5. Use default settings for all other wizard steps.
6. Launch the tool Generate MSI Baseline using Quick Launch ( ).
   - In the first step, choose where to run the tool.
   - In the "Select read mappings" step, select all control sample read mappings.
   - In the "MSI baseline options" step, click **Browse** next to the "MSI loci track" field. Navigate to CLC_References | homo_sapiens | msi_loci | msi_loci_v1.0_hg38_no_alt_analysis_set | msisensor2_loci_v1.0. Click **OK**.
   - In the next wizard step, choose to save the results.
   - In the final wizard step, choose where to save the results.
7. Once the Generate MSI Baseline tool has completed, locate the workflow in the Toolbox, right-click on the workflow and click **Open copy of workflow**.
8. Type "MSI baseline" In the "Find" field in the side panel. This will highlight the **MSI baseline** input. Double-click it and delete the text from the "Workflow role" field.
9. Click **Browse** next to the "Workflow input" field and select the MSI baseline created by the Generate MSI Baseline tool in the Navigation Area. Click **OK**.
10. Save the workflow.