User manual for
QIAGEN CLC LightSpeed Module 23.0.2

Windows, macOS and Linux

May 16, 2023

This software is for research purposes only.
## Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Introduction</td>
<td>5</td>
</tr>
<tr>
<td>1.1 System requirements</td>
<td>5</td>
</tr>
<tr>
<td>1.2 Contact information</td>
<td>5</td>
</tr>
<tr>
<td>2 Installing and uninstalling Workbench modules</td>
<td>7</td>
</tr>
<tr>
<td>2.1 Installation of modules</td>
<td>7</td>
</tr>
<tr>
<td>2.2 Licensing modules</td>
<td>8</td>
</tr>
<tr>
<td>2.3 Uninstalling modules</td>
<td>9</td>
</tr>
<tr>
<td>3 Tools</td>
<td>11</td>
</tr>
<tr>
<td>3.1 LightSpeed Methods</td>
<td>11</td>
</tr>
<tr>
<td>3.1.1 Trimming</td>
<td>11</td>
</tr>
<tr>
<td>3.1.2 Readmapping</td>
<td>12</td>
</tr>
<tr>
<td>3.1.3 Deduplication</td>
<td>12</td>
</tr>
<tr>
<td>3.1.4 Local realignment</td>
<td>12</td>
</tr>
<tr>
<td>3.1.5 Germline variant detection</td>
<td>13</td>
</tr>
<tr>
<td>3.1.6 Limitations</td>
<td>13</td>
</tr>
<tr>
<td>3.2 LightSpeed Fastq to Germline Variants</td>
<td>14</td>
</tr>
<tr>
<td>3.2.1 LightSpeed Fastq to Germline Variants outputs</td>
<td>16</td>
</tr>
<tr>
<td>3.3 Report</td>
<td>16</td>
</tr>
<tr>
<td>4 Template Workflows</td>
<td>20</td>
</tr>
<tr>
<td>4.1 Fastq to Annotated Germline Variants</td>
<td>20</td>
</tr>
<tr>
<td>4.1.1 Outputs from Fastq to Annotated Germline Variants</td>
<td>21</td>
</tr>
<tr>
<td>4.2 Fastq to Annotated Germline Variants with Coverage Analysis</td>
<td>22</td>
</tr>
<tr>
<td>4.2.1 Outputs from Fastq to Annotated Germline Variants with Coverage Analysis</td>
<td>25</td>
</tr>
</tbody>
</table>
4.3 Fastq to CNV Control ................................................................. 25
  4.3.1 Outputs from Fastq to CNV Control ................................. 27

1 Appendices ................................................................. 29

5 Licensing requirements for the QIAGEN CLC LightSpeed Module ........................................ 30
  5.1 Licensing modules on a Workbench .................................. 30
    5.1.1 Request an evaluation license .................................... 30
    Direct download ............................................................... 31
    Go to license download web page .................................... 32
    5.1.2 Download a license using a license order ID ............... 32
    Direct download ............................................................... 33
    Go to license download web page .................................... 33
    5.1.3 Import a license from a file ....................................... 33
    5.1.4 Configure License Server connection ......................... 34
    Borrowing a license .......................................................... 36
    Common issues when using a network license ...................... 37
    5.1.5 Download a static license on a non-networked computer ... 39
  5.2 Licensing Server Extensions on a CLC Server ..................... 39
    5.2.1 Download a static license on a non-networked machine ... 40
Chapter 1

Introduction

This is the manual for the QIAGEN CLC LightSpeed Module 23.0.2 - a software package providing ultra-fast secondary analysis. LightSpeed can process raw FASTQ files to germline variant calls with high accuracy without requirements for specialized hardware. This ensures flexibility in deployment solutions and cost-efficient scaling options. LightSpeed is part of our QIAGEN CLC Genomics Premium offering.

1.1 System requirements

A licensed CLC Genomics Workbench is needed to make use of the CLC LightSpeed Module. A licensed CLC Genomics Server is needed to make use of the CLC LightSpeed Server Extension.

System requirements for CLC software is provided on https://digitalinsights.qiagen.com/technical-support/system-requirements/

The system requirements for QIAGEN CLC LightSpeed Module are the same as those for other CLC Genomics Workbench, except for the following:

- All LightSpeed analyses require 32 GB RAM.
- A CPU that supports AVX2 or NEON instruction sets is required.

1.2 Contact information

QIAGEN CLC LightSpeed Module is developed by:

QIAGEN Aarhus
Silkeborgvej 2
Prismet
8000 Aarhus C
Denmark

https://digitalinsights.qiagen.com/

Email: ts-bioinformatics@qiagen.com
The QIAGEN Aarhus team continuously improves products with your interests in mind. We welcome feedback and suggestions for new features or improvements. How to contact us is described at: http://resources.qiagenbioinformatics.com/manuals/clcgenomicsworkbench/current/index.php?manual=Contact_information_citation.html.

You can also make use of our online documentation resources, including:

- **Core product manuals** https://digitalinsights.qiagen.com/technical-support/manuals/
- **Plugin manuals** https://digitalinsights.qiagen.com/products-overview/plugins/
- **Tutorials** https://digitalinsights.qiagen.com/support/tutorials/
- **Frequently Asked Questions** https://qiagen.my.salesforce-sites.com/KnowledgeBase/KnowledgeNavigatorPage
Chapter 2

Installing and uninstalling Workbench modules

The following sections describe the installation and removal of Workbench plugins and modules, on a CLC Workbench. For information about installing plugins on a CLC Genomics Server, please refer to https://resources.qiagenbioinformatics.com/manuals/clcserver/current/admin/index.php?manual=Server_plugins.html.

2.1 Installation of modules

Note: In order to install plugins and modules, the CLC Workbench must be run in administrator mode. On Windows, you can do this by right-clicking the program shortcut and choosing "Run as Administrator". On Linux and Mac, it means you must launch the program such that it is run by an administrative user.

Plugins and modules are installed and uninstalled using the Workbench Plugin Manager. To open the Plugin Manager, click on the Plugins button in the top Toolbar, or go to the menu option: Utilities | Manage Plugins…

The Plugin Manager has two tabs at the top:

- Manage Plugins An overview of your installed plugins and modules is provided under this tab.
- Download Plugins Plugins and modules available to download and install are listed in this tab.

To install a plugin, click on the Download Plugins tab (figure 2.1). Select a plugin. Information about it will be shown in the right hand panel. Click on the Download and Install button to install the plugin.

Accepting the license agreement

The End User License Agreement (EULA) must be read and accepted as part of the installation process. Please read the EULA text carefully, and if you agree to it, check the box next to the
Figure 2.1: Plugins and modules available for installation are listed in the Plugin Manager under the Download Plugins tab.

I accept these terms. If further information is requested from you, please fill this in before clicking on the Finish button.

**Installing a cpa file**

If you have a .cpa installer file for QIAGEN CLC LightSpeed Module, you can install it by clicking on the Install from File button at the bottom of the Plugin Manager.

If you are working on a system not connected to the internet, plugin and module .cpa files can be downloaded from https://digitalinsights.qiagen.com/products-overview/plugins/using a networked machine, and then transferred to the non-networked machine for installation.

**Restart to complete the installation**

Newly installed plugins and modules will be available for use after restarting the software. When you close the Plugin Manager, a dialog appears offering the opportunity to restart the CLC Workbench.

### 2.2 Licensing modules

When you have installed the QIAGEN CLC LightSpeed Module and start a tool from that module for the first time, the License Assistant will open (figure 2.2).

The License Assistant can also be launched by opening the Workbench Plugin Manager, selecting the installed module from under the Manage Plugins tab, and clicking on the button labeled Import License.

To install a license, the CLC Workbench must be run in administrator mode. On Windows, you can do this by right-clicking the program shortcut and choosing "Run as Administrator". On Linux and Mac, it means you must launch the program such that it is run by an administrative user.
The following options are available:

- **Request an evaluation license.** Request a fully functional, time-limited license.

- **Download a license.** Use the license order ID received when you purchased the software to download and install a license file.

- **Import a license from a file.** Import an existing license file, for example a file downloaded from the web-based licensing system.

- **Configure License Server connection.** If your organization has a CLC Network License Manager (or CLC License Server), select this option to configure the connection to it.


To download licenses, including evaluation licenses, your machine must have access to the external network. To install licenses on non-networked machines, please see [http://resources.qiagenbioinformatics.com/manuals/clcgenomicsworkbench/current/index.php?manual=Download_static_license_on_non_networked_machine.html](http://resources.qiagenbioinformatics.com/manuals/clcgenomicsworkbench/current/index.php?manual=Download_static_license_on_non_networked_machine.html).

### 2.3 Uninstalling modules

Plugins and modules are uninstalled using the Workbench Plugin Manager. To open the Plugin Manager, click on the **Plugins ( + )** button in the top Toolbar, or go to the menu option:

**Utilities | Manage Plugins... ( + )**

This will open the Plugin Manager (figure 2.3). Installed plugins and modules are shown under the Manage Plugins tab of the Plugins Manager.

To uninstall a plugin or module, click on its entry in the list, and click on the **Uninstall** button.

Plugins and modules are not uninstalled until the Workbench is restarted. When you close the Plugin Manager, a dialog appears offering the opportunity to restart the CLC Workbench.
Figure 2.3: Installed plugins and modules are listed in the Plugins Manager under the Manage Plugins tab.

Disabling a plugin without uninstalling it

If you do not want a plugin to be loaded the next time you start the Workbench, select it in the list under the Manage Plugins tab and click on the Disable button.
Chapter 3

Tools

The CLC LightSpeed Module contains the tool LightSpeed Fastq to Germline Variants. This chapter contains the following sections describing different aspects of the tool:

- **LightSpeed Methods** A description of each of the steps performed by the LightSpeed Fastq to Germline Variants tool.

- **LightSpeed Fastq to Germline Variants** Where to find the tool and how to run it.


3.1 LightSpeed Methods

The CLC LightSpeed Module contains the tool LightSpeed Fastq to Germline Variants which facilitates end to end NGS secondary analysis using an extensive collection of algorithms. Each individual algorithm has been optimized for short runtime with minimal memory requirements while retaining accuracy of variant detection. In the following, the overall principles of core algorithms are described.

3.1.1 Trimming

Two types of trimming are available: quality trimming and adapter trimming.

**Quality trimming** Raw reads are trimmed for low quality nucleotides. The quality limit used for trimming has been optimized for Illumina 150 bp paired end reads and cannot be adjusted.

It is possible to remove trimmed reads that are shorter than a defined threshold after quality trimming.

**Adapter trimming** The algorithm can trim adapter sequences from mapped paired-end reads. For each individual read in a pair, read sequence that extends beyond the 5’ end of the other read in the pair, is considered adapter sequence and is trimmed.

The consensus sequences for the removed R1 and R2 sequences, are included in the report.
It is possible to remove trimmed reads that are shorter than a defined threshold after adapter trimming.

### 3.1.2 Readmapping

**Indexing** When provided with a reference genome, LightSpeed first generates a Burrows-Wheeler based index of all the sequences.

**Read mapping and read pairs** LightSpeed maps reads to the indexed reference sequence. The quality scores are not stored. Single reads that are part of a paired read are mapped individually. For each read, only the most likely seeds are extended using a Needleman-Wunsch based method. LightSpeed takes the relative position of individual reads in a pair into account when estimating how likely a seed is, and prefers seeds where the distance between individual reads falls within expected distance of a paired read. Read pairs that do not map well, go through a second round of more thorough seeding. The distance at which reads can be considered as pairs, is estimated from a subset of the reads. If there is not enough data to estimate the distance, a default insert size of 1-1000 base pairs is used. Read pairs that map within the expected distance of each other are considered pairs, read pairs that map further away from each other are considered broken pairs. The algorithm has been optimized for the typical read length and error profile of Illumina 150 bp paired-end reads.

### 3.1.3 Deduplication

Deduplication can be used to collapse reads that are identical or almost identical copies and likely represent the same original DNA fragment.

Reads are deduplicated through the following steps:

1. Read pairs that are mapped in the same intervals are clustered.
2. Reads pairs that are identical or almost identical in the same cluster are considered duplicates.
3. For each group of duplicate reads, a consensus sequence is calculated.
4. The duplicate read pairs are replaced with the consensus sequence.

### 3.1.4 Local realignment

Regions where the readmapping is likely to be improved through local realignment are identified and realigned. These are generally regions where reads do not align perfectly, and the imperfect read alignments are unlikely to be caused by sequencing errors. This is for example the case where long unaligned ends (potentially representing long insertions and deletions) are present in the readmapping relative to the reference.

During local realignment, the following steps are performed for each identified region:
1. A path graph of k-mers is built. The graph contains paths corresponding to all reads as well as the reference.

2. The graph undergoes refinement where paths that are unlikely to contain a deletion or an insertion relative to the reference path are removed, and additional deletions or insertion may be added.

3. Any read that intersects the region of interest is realigned against the alignment graph.

### 3.1.5 Germline variant detection

Based on the read mapping, germline variants are identified at positions where the read alignment supports a significant difference to the reference genome.

This is achieved through a site model, where each position is first assigned a likelihood for each of the genotypes A, C, T, G, N or missing. The algorithm then iterates over the read mapping and adjusts likelihoods per position for each genotype based on observations in the data until the likelihoods no longer change. Note that broken read pairs are not considered.

Each position is then inspected, and positions where the most likely genotype(s) are different from the reference sequence are identified. As the algorithm expects the genome to be diploid and is calling germline variants, only 1 or 2 genotypes per position are considered.

**Variant types** LightSpeed Fastq to Germline Variants reports SNPs, MNVs and InDels and replacements provided that the variants are contained within at least one paired end read.

**Variant annotations** Variants identified by LightSpeed Fastq to Germline Variants are annotated with the following basic information: Chromosome, Region, Type, Reference, Allele, Reference allele, Length, Zygosity, Count, Coverage, Frequency, QUAL and Genotype. Only single base pair variants, that are not adjacent to any other variants, are assigned a QUAL score.


### 3.1.6 Limitations

**Data** LightSpeed is developed for and has been optimized on Illumina paired-end short read sequencing data. Analysis of other types of sequencing reads may not result in similar processing times or variant calls of an equivalent quality.

**Variant detection** The variant detection algorithm in LightSpeed is based on a model expecting diploid genomes. Therefore LightSpeed cannot be expected to accurately detect germline variants in genomes with other ploidies.

Alternate ploidies of sex chromosomes are not considered in the variant detection algorithm.

**Reference sequence** LightSpeed considers all chromosomes to be linear. Hence, for read mapping, circular chromosomes are linearized with position 1 starting at the junction of the chromosome. No reads will be mapped across the junction of circular chromosomes.
3.2 LightSpeed Fastq to Germline Variants

The LightSpeed Fastq to Germline Variants tool is designed to provide variant calls from raw sequencing data within a very short timeframe.

The tool can perform read trimming, mapping, deduplication, local realignment and germline variant calling. For a description of each step, see section 3.1.

LightSpeed Fastq to Germline Variants can only analyze one sample per analysis start. To analyze samples in batch, LightSpeed Fastq to Germline Variants must be included in a workflow. Template workflows for LightSpeed analysis are available (see chapter 4), but it is also possible to create custom workflows. Read about workflows here http://resources.qiagenbioinformatics.com/manuals/clgenomicsworkbench/current/index.php?manual=Workflows.html.

To run the LightSpeed tool go to:

Tools | LightSpeed | LightSpeed Fastq to Germline Variants

If you are connected to a CLC Server via your Workbench, you will be asked where you would like to run the analysis. We recommend that you run the analysis on a CLC Server when possible.

In the first wizard step, specify fastq files and a reference sequence (figure 3.1):

- **Input data**
  - **Reads (fastq)** Fastq files for analysis. At least two fastq files representing R1 and R2 reads must be provided.

- **References**
  - **References** The reference sequence that reads will be mapped to.

- **Reference masking**
  - **No masking** Reads are mapped to the full reference sequence.
  - **Exclude annotated** Reads are mapped to the full reference sequence except regions specified in the masking track.
  - **Include annotated only** Reads are only mapped to the regions specified in the masking track.
  - **Masking track** The track specifying the masking regions.

Next, options are available for trimming, read mapping and variant calling (figure 3.2):

- **Trimming**
  - **Quality trim** Reads are trimmed for low quality nucleotides.
  - **Minimum read length after quality trim** Trimmed reads shorter than this length are removed.
  - **Adapter trim** Reads are trimmed for read-through adapter sequence.
  - **Minimum read length after adapter trim** Trimmed reads shorter than this length are removed.
CHAPTER 3. TOOLS

Figure 3.1: Input fastq files and references, and, optionally, a track for reference masking.

- **Mapped read handling**
  - **Remove duplicate mapped reads** Reads likely representing PCR duplicates are collapsed.
  - **Realign mapped reads** Regions of the read mapping that may be improved through local realignment are realigned.

- **Variant detection**
  - **Restrict calling to target regions** Optional. A track defining where variants are called.
  - **Ignore non-specific matches** Reads that map equally well to more than one genomic position, are not used for variant calling.

Figure 3.2: Options for trimming, deduplication and local realignment.

In the final wizard step, choose which outputs should be generated and whether results should be saved or opened. If a reads track is selected as output, runtime will be significantly increased.
3.2.1 LightSpeed Fastq to Germline Variants outputs

LightSpeed Fastq to Germline Variants can produce the following outputs:

- **Variants track** The identified germline variants.
- **Report** A report providing information about each step, see section 3.3 for details.
- **Reads track** A read mapping. If a reads track is selected as output, runtime will be significantly increased.

3.3 Report

The LightSpeed Fastq to Germline Variants report provides information about each step that has been enabled in a given analysis. In the following, each section in the report is described.

**Summary**

- **Input read pairs** Total number of read pairs in the fastq files.
- **Read pairs removed by quality trimming** Trimmed read pairs, that after trimming are shorter than specified in the option "Minimum read length after quality trim" and have been removed.
- **Read pairs trimmed by quality trimming** Read pairs that have been trimmed and are longer than "Minimum read length after quality trim".
- **Read pairs removed by adapter trimming** Trimmed read pairs, that after trimming are shorter than specified in the option "Minimum read length after adapter trim" and have been removed.
- **Read pairs trimmed by adapter trimming** Read pairs that have been trimmed and are longer than "Minimum read length after adapter trim".
- **Non-specific mapped read pairs** Read pairs that have multiple equally good alignments to the reference.
- **Unmapped read pairs** Read pairs that did not map to the reference.
- **Mapped broken read pairs** Mapped read pairs where the distance between the individual reads in the pair exceeded the expected distance for paired reads, or where only one of the reads in the pair was mapped.
- **Removed duplicated read pairs** Read pairs that were considered PCR duplicates of other reads and were removed during deduplication.
- **Realigned regions** The number of regions that have been locally realigned.
- **Final mapped read pairs incl. non-specific** The number of mapped read pairs excluding mapped broken reads and reads removed during deduplication.
- **Final mapped read pairs excl. non-specific** The number of mapped read pairs excluding mapped broken reads, reads removed during deduplication and non-specific mapped read pairs.
CHAPTER 3. TOOLS

Quality trimming

- **Number of read pairs** Total number of read pairs in the fastq files.
- **Removed read pairs** Trimmed read pairs, that after trimming are shorter than specified in the option "Minimum read length after quality trim" and have been removed.
- **Trimmed read pairs** Read pairs that have been trimmed and are longer than "Minimum read length after quality trim".
- **Trimmed R1 reads** Trimmed R1 reads that are longer than "Minimum read length after quality trim".
- **Trimmed R2 reads** Trimmed R2 reads that are longer than "Minimum read length after quality trim".
- **Average read length before trim** Average read length of the raw reads in the fastq files.
- **Average read length after trim** Average read length after quality trimming. This read length may be longer than Average read length before trim because short reads can have been removed.

The plot Read lengths of quality trimmed reads before / after trimming shows the length and number of reads that were quality trimmed before and after trimming (figure 3.3).

![Read lengths of quality trimmed reads before / after trimming](image)

**Figure 3.3**: The number and length of quality trimmed reads before and after quality trimming.

Adapter trimming

- **Number of read pairs** Total number of read pairs in the fastq files.
- **Removed read pairs** Trimmed read pairs, that after trimming are shorter than specified in the option "Minimum read length after adapter trim" and have been removed.
- **Trimmed read pairs** Read pairs that have been trimmed and are longer than "Minimum read length after adapter trim".
- **Trimmed R1 reads** Trimmed R1 reads that are longer than "Minimum read length after adapter trim".
- **Trimmed R2 reads** Trimmed R2 reads that are longer than "Minimum read length after adapter trim".

- **Average read length before trim** Average length of the reads before adapter trimming. If quality trimming was enabled, read length after quality trim is given.

- **Average read length after trim** Average read length after adapter trimming. This read length may be longer than **Average read length before trim** because short reads can have been removed.

- **Detected R1 adapter** The consensus sequence of bases removed from R1 reads.

- **Detected R2 adapter** The consensus sequence of bases removed from R2 reads.

The plot **Read lengths of adapter trimmed reads before / after trimming** shows the number of reads as a function of read length before and after adapter trimming (figure 3.4).

![Image](image1.png)

**Figure 3.4:** The number of reads as a function of read length before and after adapter trimming.

The plot **Lengths of trimmed adapters** shows the number and lengths of trimmed adapter sequences (figure 3.5).

![Image](image2.png)

**Figure 3.5:** The number and length of trimmed adapter sequences.
Mapping statistics

- **References** The number of sequences in the reference genome.
- **Total read pairs** Total number of read pairs in the fastq files.
- **Read pairs attempted mapped** The number of read pairs left after trimming.
- **Mapped read pairs** The number of mapped read pairs.
- **Non-specific mapped read pairs** Read pairs that have multiple equally good alignments to the reference.
- **Mapped broken read pairs** Mapped read pairs where the distance between the individual reads in the pair exceeded the expected distance for paired reads, or where only one of the reads in the pair was mapped.
- **Mapped broken read pairs, one mapped** The number of broken read pairs where only one of the reads in the read pair was mapped.
- **Mapped broken read pairs, both mapped** The number of broken read pairs where both reads in the read pair were mapped.
- **Unmapped read pairs** Read pairs that did not map to the reference.
Chapter 4

Template Workflows

With the CLC LightSpeed Module the following template workflows are available:

- **Fastq to Annotated Germline Variants** identifies germline variants and annotates them with exon number and amino acid changes.

- **Fastq to Annotated Germline Variants with Coverage Analysis** identifies germline variants and annotates them with exon number and amino acid changes. A readmapping is saved, coverage metrics are calculated, and optionally, copy number variant detection is performed.

- **Fastq to CNV Control** generates control coverage tables for copy number variant detection.

4.1 Fastq to Annotated Germline Variants

The **Fastq to Annotated Germline Variants** template workflow identifies germline variants and annotates these with exon number and amino acid changes.

The workflow can be used to identify and annotate variants in both targeted sequencing and whole genome sequencing pipelines.

The workflow can be found at: [Template Workflows | LightSpeed Workflows](#) | Fastq to Annotated Germline Variants

If you are connected to a CLC Server via your Workbench, you will be asked where you would like to run the analysis. We recommend that you run the analysis on a CLC Server when possible.

In the first wizard step, select a Reference Data Set (figure 4.1). If you have not downloaded the Reference Data Set yet, the dialog will suggest the relevant data set and offer the opportunity to download it using the Download to Workbench button.


In the Fastq to Germline Variants wizard step (figure 4.2) you have the following options:

- **Fastq reads** Press **Browse** to select fastq files for analysis.
• **Remove duplicate mapped reads** Duplicate mapped reads are per default replaced with a consensus read. Untick if duplicate mapped reads should be retained. See section 3.1.3 for additional details.

• **Restrict calling to target regions** Optional, if a targeted protocol is used, provide target regions here.

• **Batch** Select if fastq files from different samples are used as input, and each sample should be analyzed individually. LightSpeed supports analysis of Illumina paired-end sequencing reads in fastq format. The names of the fastq files must follow standard Illumina naming scheme to allow the tool to identify individual fastq files as belonging to the same sample.

• **Join lanes when batching** Select to join fastq files from the same sample that were sequenced on different lanes.

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

### 4.1.1 Outputs from Fastq to Annotated Germline Variants

The **Fastq to Annotated Germline Variants** template workflow produces the following outputs:

• **Germline Variants** The variant track with the annotated variants.

• **LightSpeed Report** A report summarizing details of each analysis step performed by the LightSpeed Fastq to Germline Variants tool.
4.2 Fastq to Annotated Germline Variants with Coverage Analysis

The **Fastq to Annotated Germline Variants with Coverage Analysis** template workflow:

- Identifies germline variants and annotates these with exon number and amino acid changes.
- Produces a read mapping.
- Reports coverage at target and gene level.
- Optionally identifies copy number variants (CNVs).

The workflow can only be used with targeted data.

The runtime of this workflow is significantly longer than the runtime of **Fast to Annotated Germline Variants** (section 4.1), because a read mapping track is saved.

Fastq to Annotated Germline Variants with Coverage Analysis can be found at:

[Template Workflows](#) | [LightSpeed Workflows](#) | [Fastq to Annotated Germline Variants with Coverage Analysis](#)

If you are connected to a CLC Server via your Workbench, you will be asked where you would like to run the analysis. We recommend that you run the analysis on a CLC Server when possible.

In the first wizard step, select the target regions (figure 4.3).

Next, select a Reference Data Set (figure 4.4). If you have not downloaded the Reference Data Set yet, the dialog will suggest the relevant data set and offer the opportunity to download it using the Download to Workbench button.
Figure 4.3: Select the target regions.

If none of the available reference data sets are appropriate, custom reference data sets can be created, see http://resources.qiagenbioinformatics.com/manuals/clcgenomicsworkbench/current/index.php?manual=Custom_Sets.html.

Figure 4.4: Select a reference data set.

In the LightSpeed Fastq to Germline Variants with Coverage Analysis wizard step (figure 4.5) you have the following options:

- **Fastq reads** Press **Browse** to select fastq files for analysis.
- **Remove duplicate mapped reads** Duplicate mapped reads are per default replaced with a consensus read. Untick if duplicate mapped reads should be retained. See section 3.1.3 for additional details.
• **Batch** Select if fastq files from different samples are used as input, and each sample should be analyzed individually. LightSpeed supports analysis of Illumina paired-end sequencing reads in fastq format. The names of the fastq files must follow standard Illumina naming scheme to allow the tool to identify individual fastq files as belonging to the same sample.

• **Join lanes when batching** Select to join fastq files from the same sample that were analyzed on different lanes.

**Figure 4.5: Select fastq files.**

In the wizard step Copy Number Variant Detection (CNVs), it is possible to specify control coverage tables or read mappings for copy number variant detection (figure 4.6). If controls are not provided, copy number variant detection will not be performed. Read about copy number variant detection here [http://resources.qiagenbioinformatics.com/manuals/clcgenomicsworkbench/current/index.php?manual=Copy_Number_Variant_Detection.html](http://resources.qiagenbioinformatics.com/manuals/clcgenomicsworkbench/current/index.php?manual=Copy_Number_Variant_Detection.html).

Note that for CNV detection it is important that the same processing is applied to control samples and the sample that is tested for CNVs. We recommend using the LightSpeed template workflow **Fastq to CNV Control** to create appropriate control coverage tables, see section 4.3.

**Figure 4.6: Select control coverage tables or read mappings for copy number variant detection.**

In the final wizard step, choose to **Save** the results of the workflow and specify a location in the Navigation Area before clicking **Finish**.
4.2.1 Outputs from Fastq to Annotated Germline Variants with Coverage Analysis

The Fastq to Annotated Germline Variants with Coverage Analysis template workflow produces the following outputs:

- **Germline Variants** The variant track with the annotated variants.
- **LightSpeed Report** A report summarizing details of each analysis step performed by the LightSpeed Fastq to Germline Variants tool.
- **Genome Browser View** A track list containing the Germline Variants, the Amino Acid Track, the Target Region Statistics Track, the Gene-level CNV Track as well as the Reference sequence and the Genes, mRNA and CDS tracks.
- **Read Mapping** A read mapping track.
- **Amino Acid Track** A track providing a graphical representation of identified amino acid changes.
- **CNV Results Report** A report providing an overview of identified CNVs.
- **Target-level CNV Track** An annotation track providing CNV results per target.
- **Gene-level CNV Track** An annotation track providing CNV results per gene.
- **Region-level CNV Track** An annotation track providing CNV results per region, where regions are formed from adjacent targets with similar CNV states.
- **Coverage Report** A report summarizing coverage.
- **Target Region Statistics Track** A track providing coverage information per target region.
- **Gene Coverage Track** A track providing coverage information per gene.

The Amino Acid Track is produced by Amino Acid Changes.

The CNV Results Report, and the Target, Gene and Region-level CNV Tracks are produced by Copy Number Variant Detection (CNVs).

The Coverage Report, Target Region Statistics Track and the Gene Coverage Track are produced by QC for Targeted Sequencing.

4.3 Fastq to CNV Control

The Fastq to CNV Control template workflow produces coverage tables that can be used as controls for copy number variant detection.

The workflow can only be used with targeted data.
Use the workflow to generate coverage tables for the **Fastq to Annotated Germline Variants with Coverage Analysis** template workflow (section 4.2).

**Fastq to CNV Control** can be found at:

[Template Workflows | LightSpeed Workflows](#) | Fastq to CNV Control

If you are connected to a CLC Server via your Workbench, you will be asked where you would like to run the analysis. We recommend that you run the analysis on a CLC Server when possible.

In the first wizard step, select the target regions (figure 4.7).

The target regions must be identical to the target regions that will later be used for copy number variant detection together with the control coverage tables.

![Figure 4.7: Select the target regions.](#)

Next, select a Reference Data Set (figure 4.8). If you have not downloaded the Reference Data Set yet, the dialog will suggest the relevant data set and offer the opportunity to download it using the Download to Workbench button.


In the LightSpeed Fastq to Germline Variants wizard step (figure 4.9) you have the following options:

- **Fastq reads** Press **Browse** to select fastq files for analysis.
- **Remove duplicate mapped reads** Duplicate mapped reads are per default replaced with a consensus read. Untick if duplicate mapped reads should be retained. See section 3.1.3 for additional details.
- **Batch** Select if fastq files from different samples are used as input, and each sample should be analyzed individually. LightSpeed supports analysis of Illumina paired-end sequencing reads in fastq format. The names of the fastq files must follow standard Illumina naming scheme to allow the tool to identify individual fastq files as belonging to the same sample.
- **Join lanes when batching** Select to join fastq files from the same sample that were analyzed on different lanes.
In the final wizard step, choose to Save the results of the workflow and specify a location in the Navigation Area before clicking Finish.

### 4.3.1 Outputs from Fastq to CNV Control

The Fastq to CNV Control template workflow produces the following outputs:

- **Coverage Table** A table providing coverage information per position in the target regions.

  The coverage table can be used as control for copy number variant detection, either in the template workflow Fastq to Annotated Germline Variants with Coverage Analysis (section 4.2), or directly in the tool Copy Number Variant Detection (CNVs) [http://resources.](http://resources.)
• **LightSpeed Report** A report summarizing details of each analysis step performed by the LightSpeed Fastq to Germline Variants tool.

• **Coverage Report** A report summarizing coverage.

• **Target Region Statistics Track** A track providing coverage information per target region.

The **Coverage Table**, **Coverage Report**, and the **Target Region Statistics Track** are produced by **QC for Targeted Sequencing**.
Part I

Appendices
Chapter 5

Licensing requirements for the QIAGEN CLC LightSpeed Module

To use tools delivered by the QIAGEN CLC LightSpeed Module or CLC LightSpeed Server Extension, you will need a license. This chapter describes how to install and configure such licenses on a CLC Workbench or CLC Server. QIAGEN CLC LightSpeed Module can be installed in Viewing Mode on a CLC Workbench to access data created using the module’s tools without requiring a license.

5.1 Licensing modules on a Workbench

The License Assistant can be launched by opening the Workbench Plugin Manager, selecting the installed module from under the Manage Plugins tab, and clicking on the button labeled Import License.

To install a license, the CLC Workbench must be run in administrator mode. On Windows, you can do this by right-clicking the program shortcut and choosing "Run as Administrator". On Linux and Mac, it means you must launch the program such that it is run by an administrative user.

You need a license...
In order to load the plugin "CLC Genome Finishing Module" you need a valid license.
Please choose how you would like to obtain a license for this plugin.

- Request an evaluation license
  Choose this option if you would like to try out the plugin for 14 days.
  Please note that only one evaluation license will be allowed for each computer.

- Download a license
  Use a license order ID to download a static license.

- Import a license from a file
  Import a static license from an existing license file.

- Configure License Server connection
  Configure the necessary connection for the software to connect to a CLC License Server that hosts network license(s) for the product. This option also allows you to edit or delete an existing configuration.

Figure 5.1: The License Assistant provides options for licensing modules installed on the Workbench.

5.1.1 Request an evaluation license

We offer a fully functional version of the QIAGEN CLC LightSpeed Module free of charge for a 14 day period for evaluation purposes. The 14 day period commences when the evaluation license
is downloaded. If you have questions about QIAGEN CLC LightSpeed Module features or product licensing options, please send an email to bioinformaticssales@qiagen.com.

When you choose the option **Request an evaluation license**, the dialog shown in figure 5.2 opens.

![Request an evaluation license](image)

**Figure 5.2: Choosing between direct download or going to the license download web page.**

In this dialog, there are two options:

- **Direct Download.** Download the license directly. This method requires that the Workbench has access to the external network.

- **Go to CLC License Download web page.** The online license download form will be opened in a web browser. This option is suitable for when downloading a license for use on another machine that does not have access to the external network, and thus cannot access the QIAGEN Aarhus servers.

After selecting your method of choice, click on the button labeled **Next**.

**Direct download**

After choosing the **Direct Download** option and clicking on the button labeled **Next**, a dialog similar to that shown in figure 5.3 will appear if the license is successfully downloaded and installed.

![Requesting a license](image)

**Figure 5.3: A license has been successfully downloaded and installed for use.**

When the license has been downloaded and installed, the **Next** button will be enabled.

If there is a problem, a dialog will appear indicating this.
Go to license download web page

After choosing the Go to CLC License Download web page option and clicking on the button labeled Next, the license download form will be opened in a web browser, as shown in figure 5.4.

Click on the Download License button and then save the license file.

Back in the Workbench window, you will now see the dialog shown in 5.5.

Click on the Choose License File button, find the saved license file and select it. Then click on the Next button.

5.1.2 Download a license using a license order ID

Using a license order ID, you can download a license file via the Workbench or using an online form. When you have chosen this option and clicked Next button, you will see the dialog shown in 5.6. Enter your license order ID into the text field under the title License Order-ID. (The ID can be pasted into the box after copying it and then using menus or key combinations like Ctrl+V on some system or ⌘ + V on Mac).

In this dialog, there are two options:

- Direct Download. Download the license directly. This method requires that the Workbench has access to the external network.
- Go to CLC License Download web page. The online license download form will be opened in a web browser. This option is suitable for when downloading a license for use on another machine that does not have access to the external network, and thus cannot access the QIAGEN Aarhus servers.
After selecting your method of choice, click on the button labeled **Next**.

**Direct download**

After choosing the **Direct Download** option and clicking on the button labeled **Next**, a dialog similar to that shown in figure 5.7 will appear if the license is successfully downloaded and installed.

When the license has been downloaded and installed, the **Next** button will be enabled.

If there is a problem, a dialog will appear indicating this.

**Go to license download web page**

After choosing the **Go to CLC License Download web page** option and clicking on the button labeled **Next**, the license download form will be opened in a web browser, as shown in figure 5.8.

Click on the **Download License** button and then save the license file.

Back in the Workbench window, you will now see the dialog shown in 5.9.

Click on the **Choose License File** button, find the saved license file and select it. Then click on the **Next** button.

**5.1.3 Import a license from a file**

If you already have a license file associated with the host ID of your machine, it can be imported using this option.
When you have clicked on the **Next** button, you will see the dialog shown in **5.10**.

Click the **Choose License File** button and browse to find the license file. When you have selected the file, click on the **Next** button.

### 5.1.4 Configure License Server connection

If your organization is running a **CLC Network License Manager** or CLC License Server, you can configure your Workbench to connect to it to get a license for the module.

To configure the Workbench to connect to a **CLC Network License Manager** or CLC License Server, select the **Configure License Server connection** option and click on the **Next** button. A dialog
appears, as shown in figure 5.11.

Figure 5.11: Connecting to a CLC Network License Manager or CLC License Server.

The options in that dialog are:

- **Enable license manager connection.** This box must be checked for the Workbench is to contact the CLC Network License Manager or CLC License Server to get a license for the CLC Workbench.

- **Automatically detect license manager.** By checking this option the Workbench will look for a CLC Network License Manager or CLC License Server accessible from the Workbench. Automatic server discovery sends UDP broadcasts from the Workbench on port 6200. Available license servers respond to the broadcast. The Workbench then uses TCP communication for to get a license, if one is available. Automatic server discovery works only on local networks and will not work on WAN or VPN connections. Automatic server discovery is not guaranteed to work on all networks. If you are working on an enterprise network on where local firewalls or routers cut off UDP broadcast traffic, then you may need to configure the details of the CLC Network License Manager or CLC License Server using the **Manually specify license manager** option instead.

- **Manually specify license manager.** Select this option to enter the details of the machine the CLC Network License Manager or CLC License Server software is running on, specifically:
  - **Host name.** The address of the machine the CLC Network License Manager or CLC License Server software is running on.
  - **Port.** The port used by the CLC Network License Manager or CLC License Server to receive requests.

- **Use custom username when requesting a license.** Optional. If this is checked, a username can be entered that will be used when requesting a network license instead of the username of the account being used to run the Workbench.

- **Disable license borrowing on this computer.** Check this box if you do not want users of the computer to borrow a license. See section 5.1.4 for further details.
Special note on modules needing a license

A valid module license is needed to start a module tool, or a workflow including a module tool. Network licenses for modules are valid for four hours after starting the tool or the workflow. A process started (whether a module tool or a workflow including a module tool) will always be completed, even if its completion exceeds the four hours period where the license is valid.

If the tool or the workflow completes before the four hour validity period, it is possible to start a new tool or a workflow, and this will always refresh the validity of the license to a full four hours period. However, if the tool or the workflow completes after the four hour validity period, a new license will need to be requested after that to start the next tool or workflow.

These measures ensure that more licenses are available to active users, rather than blocked on an inactive computer, i.e., where the workbench would be open but not in use.

Borrowing a license

A CLC Workbench using a network license normally needs to maintain a connection to the CLC Network License Manager or CLC License Server. However, if allowed by the network license administrator, network licenses can be borrowed for offline use. During the period a license has been borrowed, there will be one less network license available for other users.

If administrator has chosen not to allow the borrowing of network licenses, then the information in this section is not relevant.

The Workbench must be connected to the CLC Network License Manager or CLC License Server at the point when the license is borrowed. The procedure for borrowing a license is:

1. Go to the Workbench menu option: Help | License Manager
2. Click on the "Borrow License" tab to display the dialog shown in figure 5.12.
3. Select the license(s) that you wish to borrow by clicking in the checkboxes in the Borrow column in the License overview panel.
4. Choose the length of time you wish to borrow the license(s) for using the drop down list in the Borrow License tab. By default the maximum is 7 days, but network license administrators can specify a lower limit than this.
5. Click Borrow Selected Licenses.
6. Close the License Manager when you are done.

You can now go offline and continue working with the CLC Workbench. When the time period you borrowed the license for has elapsed, the network license will be again made available for other users. To continue using CLC Workbench with a license, you will need to connect to the network again so the Workbench can request another license.

You can return borrowed licenses early if you wish by started up the License Manager, opening the "Borrow License" tab, and clicking on the Return Borrowed Licenses button.
Common issues when using a network license

- No license available at the moment If all licenses are in use, you will see a dialog like that shown in figure 5.13 when you start up the Workbench.

You will need to wait for at least one license to be returned before you can continue to work with a fully functional copy of the software. If running out of licenses is a frequent issue, you may wish to discuss this with your administrator.

Clicking on the Viewing Mode button in the dialog allows you to run the CLC Workbench for viewing data, and for basic analyses, import and export.

- Lost connection to the CLC Network License Manager or CLC License Server If the Workbench connection to the CLC Network License Manager of CLC License Server is lost, you will see a dialog like that shown in figure 5.14.

If you have chosen the option to Automatically detect license manager and you have not succeeded in connecting to the CLC Network License Manager or CLC License Server before, please check with your local IT support that automatic detection will be possible to do at your site. If it is not, you will need to specify the settings, as described earlier in this section.
CHAPTER 5. LICENSING REQUIREMENTS FOR THE QIAGEN CLC LIGHTSPEED MODULE

Figure 5.14: Here, the Workbench is unable to establish a connection to obtain a network license.

If you have successfully contacted the CLC Network License Manager or CLC License Server from your Workbench previously, please consider discussing this issue with your administrator, for example, making sure that the CLC Network License Manager or CLC License Server is running and that your Workbench is able to connect to it.

There may be situations where you wish to use a different license or view information about the license(s) the Workbench is currently using. To do this, open the License Manager using the menu option:

Help | License Manager

The license manager is shown in figure 5.15.

This dialog can be used to:

• See information about the license (e.g. what kind of license, when it expires)
• Configure how to connect to a license server (Configure License Server the button at the lower left corner). Clicking this button will display a dialog similar to figure 5.11.
• Upgrade from an evaluation license by clicking the Upgrade license button. This will display the dialog shown in figure 2.2.
• Export license information to a text file.
• Borrow a license

If you wish to switch away from using a network license, click on the button to Configure License Server and uncheck the box beside the text Enable license server connection in the dialog. When you restart the Workbench, you can set up the new license as described in section 2.2.

5.1.5 Download a static license on a non-networked computer

To download a static license for a machine that does not have direct access to the external network, you can follow the steps below:

• Install the QIAGEN CLC LightSpeed Module on the machine you wish to run the software on.
• Start up the software as an administrative user and find the host ID of the machine that you will run the CLC Workbench on. You can see the host ID the machine reported at the bottom of the License Manager window in grey text.
• Make a copy of this host ID such that you can use it on a machine that has internet access.
• Go to a computer with internet access, open a browser window and go to the network license download web page:
  https://secure.clcbio.com/LmxWSv3/GetLicenseFile
• Paste in your license order ID and the host ID that you noted down in the relevant boxes on the webpage.
• Click ‘download license’ and save the resulting .lic file.
• Open the Workbench on your non-networked machine. In the Workbench license manager choose ‘Import a license from a file’. In the resulting dialog click ‘choose license file’ to browse the location of the .lic file you have just downloaded.
  If the License Manager does not start up by default, you can start it up by going to the Help menu and choosing License Manager.
• Click on the Next button and go through the remaining steps of the license manager wizard.

5.2 Licensing Server Extensions on a CLC Server

Licenses for Server Extensions are downloaded and installed by a server administrator using the CLC Server web administrative interface.

Licenses are installed on a single server or on the master node of a job node or grid node setup.

To download and install a license:
• Log into the web administrative interface of the single server or master node as an administrative user.

• Under the Management ( ) tab, open the Download License ( ) tab.

• Enter the Order ID supplied by QIAGEN into the Order ID field and click on the “Download and Install License...” button (figure 5.16).

Please contact ts-bioinformatics@qiagen.com if you have not received an Order ID.

The CLC Server must be restarted for new license files to be loaded. Details about restarting can be found at resources.qiagenbioinformatics.com/manuals/clcserver/current/admin/index.php?manual=Starting_stopping_server.html.

Each time you download a license file, a new file is created in the licenses folder under the CLC Server installation area. If you are upgrading an existing license file, delete the old file from this area before restarting.

If you are working on a system that does not have access to the external network, then please refer to section 5.2.1.

Figure 5.16: License management in done under the Management tab tab.

### 5.2.1 Download a static license on a non-networked machine

Follow the steps below to download a static license for a server machine that does not have direct access to the external network.

• Determine the host ID of the machine the server software will be running on. This can be done by running the back-end license download tool, which prints the host ID of the system to the terminal. The download tool is located in the installation folder of the CLC Server. The tool name depends on the system you are working on:
  - Linux: downloadlicense
  - Mac: downloadlicense.command
  - Windows: licensedownload.bat

In the case of a job or grid node setup, the host ID should be for the machine that will act as the CLC Server master node.
• Make a copy of this host ID such that you can use it on a machine that has internet access.

• Go to a computer with internet access, open a browser window and go to the license download web page:
  https://secure.clcbio.com/LmxWSv3/GetLicenseFile

• Paste in your license order ID and the host ID that you noted down earlier into the relevant boxes on the webpage.

• Click on 'download license' and save the resulting .lic file.

• On the machine with the host ID you specified when downloading the license file, place the license file in the folder called 'licenses' in the CLC Server installation directory.

• Restart the CLC software.