Improve Reads with Trimmomatic Element

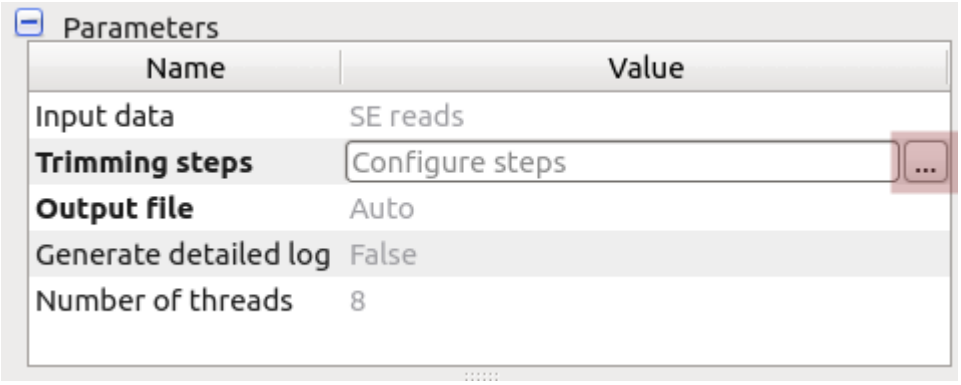
Trimmomatic is a fast, multithreaded command line tool that can be used to trim and crop Illumina (FASTQ) data as well as to remove adapters.

Element type: trimmomatic

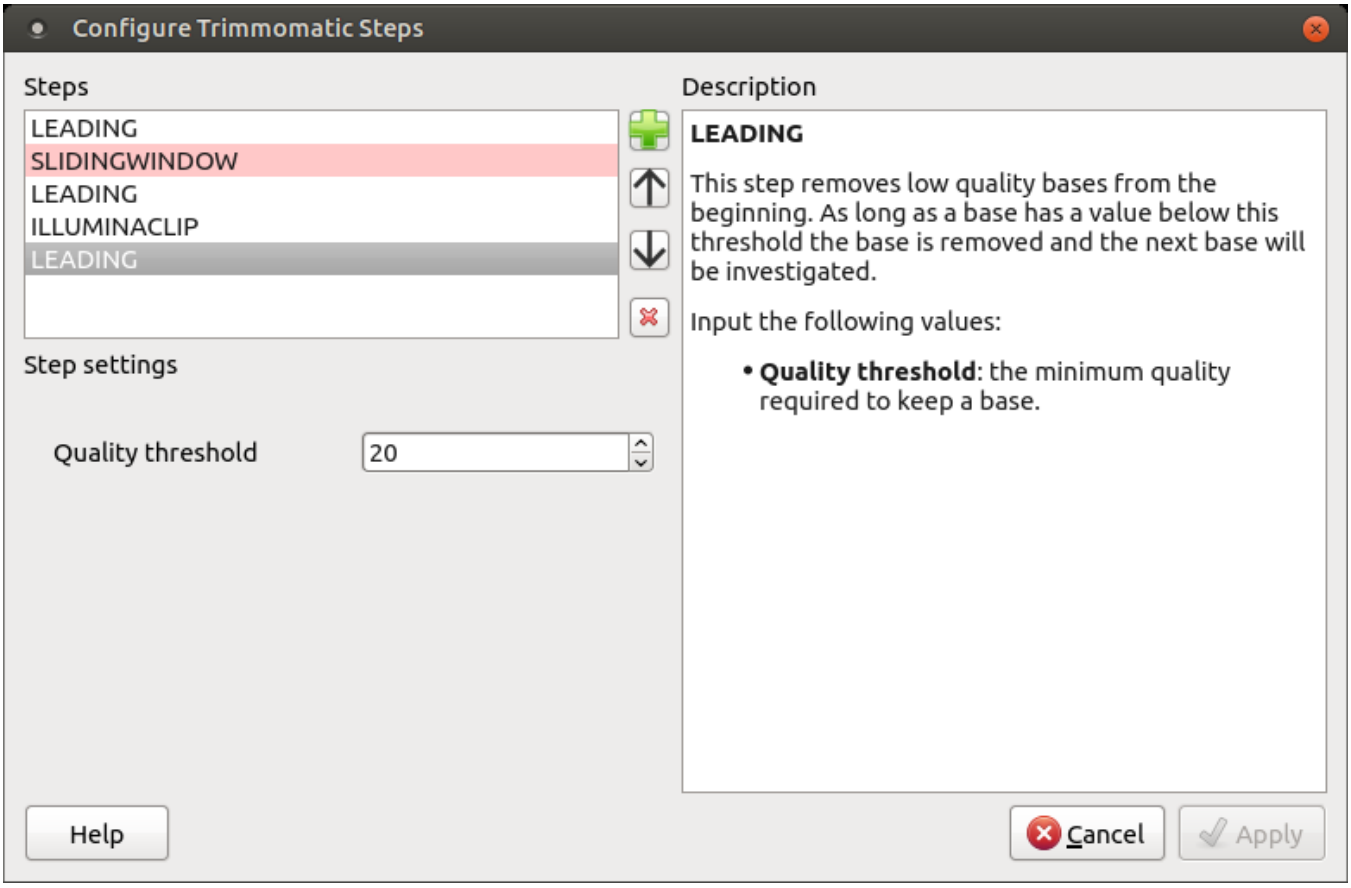
Parameters

Parameter	Description	Default value	Parameter in Workflow File	Type
Input data	Set the type of the input reads: single-end (SE) or paired-end (PE). One or two slots of the input port are used depending on the value of the parameter. Pass URL(s) to data to these slots. Note that the paired-end mode will use additional information contained in paired reads to better find an adapter or PCR primer fragments introduced by the library preparation process.	SE reads	**input-data**	*string*
Trimming steps	Configure trimming steps that should be performed by Trimmomatic.	configure steps	**trimming-steps**	*string*
Output file	Specify the output file name.	auto	**output-url**	*string*
Generate detailed log	Select "True"to generate a file with log of all read trimmings, indicating the following details (-trimlog): • thread name • the surviving sequence length • the location of the first surviving base, aka. the amount trimmed from the start • the location of the last surviving base in the original read • the amount trimmed from the end	False	**generate-log**	*bool*
Number of threads	Use multiple threads (-threads).	8	**threads**	*numeric*

To configure trimming steps use the following button:



The following dialog will appear:



Click the *Add new* step button and select a step. The following options are available:

- ILLUMINACLIP: Cutadapterandotherillumina-specific sequences from the read.
- SLIDINGWINDOW: Perform a sliding window trimming, cutting once the average quality within the window falls below a threshold.
- LEADING: Cut bases off the start of a read, if below a threshold quality.
- TRAILING: Cut bases off the end of a read, if below a threshold quality.
- CROP: Cut the read to a specified length.
- HEADCROP: Cut the specified number of bases from the start of the read.
- MINLEN: Drop the read if it is below a specified length.
- AVGQUAL: Drop the read if the average quality is below the specified level.
- TOPHRED33: Convert quality scores to Phred-33.
- TOPHRED64: Convert quality scores to Phred-64.

Each step has the own parameters:

AVGQUAL

This step drops a read if the average quality is below the specified level.

Input the following values:

- Quality threshold: the minimum average quality required to keep a read.

CROP

This step removes bases regardless of quality from the end of thread, so that the readhas maximally the specified length after this step has been performed. Steps performed after CROP might of course further shorten the read.

Input the following values:

- Length: the number of bases to keep, from the start of the read.

HEADCROP

This step removes the specified number of bases, regardless of quality, from the beginning of the read.

Input the following values:

- Length: the number of bases to remove from the start of the read.

ILLUMINACLIP

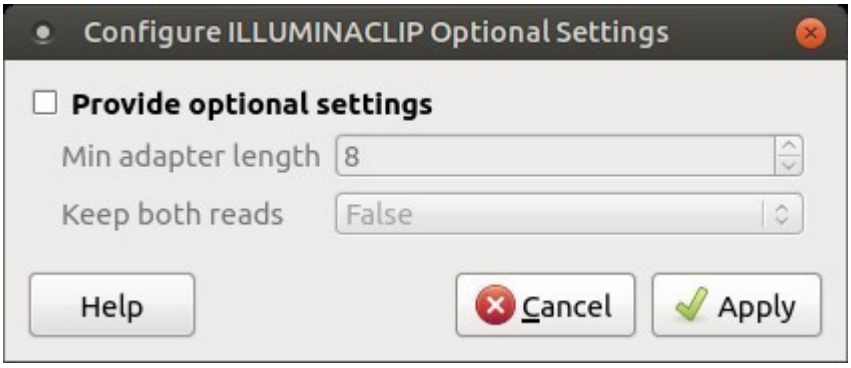
This step is used to find and remove Illumina adapters.

Trimmomatic first compares short sections of an adapter and a read. If they match enough, the entire alignment between the read and adapter is scored. For paired-end reads, the "palindrome" approach is also used to improve the result. See Trimmomatic manual for details.

Input the following values:

- Adapter sequences: a FASTA file with the adapter sequences. Files for TruSeq2 (GAII machines), TruSeq3 (HiSeq and MiSeq machines) and Nextera kits for SE and PE reads are now available by default. The naming of the various sequences within the specified file determines how they are used.
- Seed mismatches: the maximum mismatch count in short sections which will still allow a full match to be performed.
- Simple clip threshold: a threshold for simple alignment mode. Values between 7 and 15 are recommended. A perfect match of a 12 base sequence will score just over 7, while 25 bases are needed to score 15.
- Palindrome clip threshold: a threshold for palindrome alignment mode. For palindromic matches, a longer alignment is possible. Therefore the threshold can be in the range of 30. Even though this threshold is very high (requiring a match of almost 50 bases) Trimmomatic is still able to identify very, very short adapter fragments.

There are also two optional parameters for palindrome mode: Min adapter length and Keep both reads. Use the following dialog. To call the dialog press the *Optional* button.



LEADING

This step removes low-quality bases from the beginning. As long as a base has a value below this threshold the base is removed and the next base will be investigated.

Input the following values:

- Quality threshold: the minimum quality required to keep a base.

MAXINFO

This step performs an adaptive quality trim, balancing the benefits of retaining longer reads against the costs of retaining bases with errors. See Trimmomatic manual for details.

Input the following values:

- Target length: the read length which is likely to allow the location of the read within the target sequence. Extremely short reads, which can be placed into many different locations, provide little value. Typically, the length would be in the order of 40 bases, however, the value also depends on the size and complexity of the target sequence.
- Strictness: the balance between preserving as much read length as possible vs. removal of incorrect bases. A low value of this parameter (0.8) favours read correctness.

MINLEN

This step removes reads that fall below the specified minimum length. If required, it should normally be after all other processing steps. Reads removed by this step will be counted and included in the "dropped reads" count.

Input the following values:

- Length: the minimum length of reads to be kept.

SLIDINGWINDOW

This step performs a sliding window trimming, cutting once the average quality within the window falls below a threshold. By considering multiple bases, a single poor quality base will not cause the removal of high-quality data later in the read.

Input the following values:

- Window size: the number of bases to an average across.
- Quality threshold: the average quality required.

TOPHRED33

This step (re)encodes the quality part of the FASTQ file to base 33.

TOPHRED64

This step (re)encodes the quality part of the FASTQ file to base 64.

TRAILING

This step removes low-quality bases from the end. As long as a base has a value below this threshold the base is removed and the next base (i.e. the preceding one) will be investigated. This approach can be used removing the special Illumina " low-quality segment" regions (which are marked with a quality score of 2), but SLIDINGWINDOW or MAXINFO are recommended instead.

Input the following values:

- Quality threshold: the minimum quality required to keep a base.

To remove a step use the *Remove selected step* button. The pink highlighting means the required parameter has not been set.

Input/Output Ports

The element has 1 *input port*:

Name in GUI: Input FASTQ file(s)

Name in Workflow File: in

Slots:

Slot In GUI	Slot in Workflow File	Type
Input FASTQ URL	**reads-url1**	*string*

And 1 *output port*:

Name in GUI: Improved FASTQ file(s)

Name in Workflow File: out-file

Slots:

Slot In GUI	Slot in Workflow File	Type
Output FASTQ URL	**reads-url1**	*string*