Raw DNA-Seq Processing

> ⓘ Download and install the UGENE FULL or NGS package to use this pipeline.

Use this workflow sample to process raw DNA-seq next-generation sequencing (NGS) data from the Illumina platform. The processing includes:

- *Filtration:*
    - Filtering of the NGS short reads by the CASAVA 1.8 header;
    - Trimming of the short reads by quality;
- *Mapping:*
    - Mapping of the short reads to the specified reference sequence (the BWA-MEM tool is used in the sample);
- *Post-filtration:*
    - Filtering of the aligned short reads by SAMtools to remove reads with low mapping quality, unpaired/unaligned reads;
    - Removing of duplicated short reads.

The result filtered short reads assembly is provided in the SAM format. Intermediate data files are also available in the output.

> ✓ **How to Use This Sample**
>
> If you haven't used the workflow samples in UGENE before, look at the "How to Use Sample Workflows" section of the documentation.

Workflow Sample Location

The workflow sample "Raw DNA-Seq processing" can be found in the "NGS" section of the Workflow Designer samples.

Workflow Image

There are two versions of the workflow available. The workflow for single-end reads looks as follows:
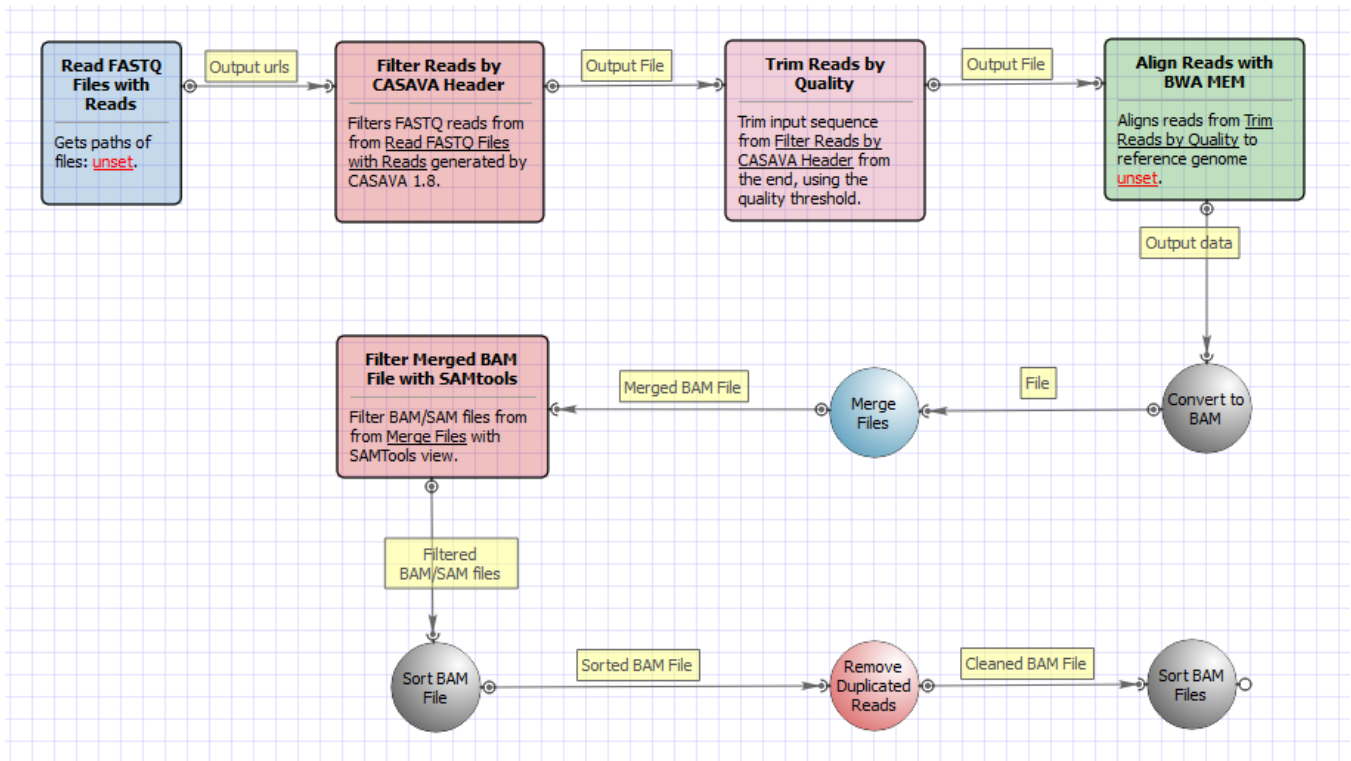
The workflow for paired-end short appearance is the following:

Read FASTQ Files with Reads 1

Gets paths of files: unset.

Output urls →

Filter Reads by CASAVA Header

Filters FASTQ reads from from Read FASTQ Files with Reads 1 generated by CASAVA 1.8.

Output File →

Trim Reads by Quality

Trim input sequence from Filter Reads by CASAVA Header from the end, using the quality threshold.

Output File →

Align Reads with BWA MEM

Aligns upstream oriented reads from Trim Reads by Quality and downstream oriented reads from Trim Reads by Quality to reference genome unset.

Read FASTQ Files with Reads 2

Gets paths of files: unset.

Output urls →

Filter Reads by CASAVA Header

Filters FASTQ reads from from Read FASTQ Files with Reads 2 generated by CASAVA 1.8.

Output File →

Trim Reads by Quality

Trim input sequence from Filter Reads by CASAVA Header from the end, using the quality threshold.

Output File →

Output data

Filter Merged BAM File with SAMtools

Filter BAM/SAM files from from Merge Files with SAMTools view.

Merged BAM File ← **Merge Files** ← **File** ← **Convert to BAM**

Filtered BAM/SAM files

Sort BAM File → **Sorted BAM File** → **Remove Duplicated Reads** → **Cleaned BAM File** → **Sort BAM File**