## **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:



Workflow Wizard

The wizard for single-end reads has 5 page.

1. Input data: On this page you must input FASTQ file(s).

Raw DNA-seq Processing Wizard		? X
	Short Reads FASTQ files Required	
Defaults	Next >	Cancel

2. Pre-processing: On this page you can modify filtration parameters.

Raw DNA-seq Processing Wiza	rd	2	X
	Pre-prov Filtration Quality thresho Min length Adapters	Cessing 1 1 yrk/ugene/data/adapters/adapters.fasta	•
U gene Defaults		< Back Next > Cance	21

3. <u>Mapping:</u> On this page you must input reference and optionally modify advanced parameters.

Raw DNA-seq Processing Wiza	rd ? X
	Mapping BWA-MEM Mapping Reference genor Required Number of threa 8
Defaults	< Back Next > Cancel

4. Post-processing: On this page you can modify post-processing parameters.

Raw DNA-seq Processing Wiza	rd ? X
	Post-processing Filtration MAPQ threshold 1 Skip flag The read is unmapped • Region Remove Duplicates For single-end re True •
Defaults	< Back Next > Cancel

5. <u>Output data:</u> On this page you must input output parameters.

Raw DNA-seq Processing Wiza	rd ? X
	Output data Aligned data Output file name out.sam Output directory output Filtered FASTQ Show filtered fastqparameters +
Defaults	< Back Apply Run Cancel

The wizard for paired-end reads has 5 page.

1. Input data: On this page you must input FASTQ file(s).

Raw DNA-seq Processing Wiza	rd			2	X
	Input da	ata			
	Short Reads				
	FASTQ files wit	Required			
Defaults			Next >	Can	cel

2. <u>Pre-processing:</u> On this page you can modify filtration parameters.

Raw DNA-seq Processing Wiza	rd Pre-proc Filtration Base quality Reads length Base quality for Reads length for	20 1 20 1	
	Base quality for Reads length for Adapters Adapters for pai	20 1 ork/ugene/data/adapters/adapters.fas	↓   ↓   ↓   ↓
Defaults		< Back Next > Ca	ncel

3. <u>Mapping:</u> On this page you must input reference and optionally modify advanced parameters.

Raw DNA-seq Processing Wiza	rd P X
	Mapping BWA-MEM Mapping Reference genor Required Number of threa 8
Defaults	< Back Next > Cancel

4. Post-processing: On this page you can modify post-processing parameters.

Raw DNA-seq Processing Wiza	rd ? X
	Post-processing Filtration MAPQ threshold 1 Skip flag The read is unmapped • Region Remove Duplicates For single-end retrue •
Defaults	< Back Next > Cancel

5. Output data: On this page you must input output parameters.

Raw DNA-seq Processing Wiza	rd ? X
	Output data Aligned data Output file name out.sam Output directory ouput Filtered FASTQ Show filtered fastqparameters +
Defaults	< Back Apply Run Cancel