

# Map RNA-Seq Reads with TopHat Element

TopHat is a program for mapping RNA-Seq reads to a long reference sequence. It uses Bowtie or Bowtie2 to map the reads and then analyzes the mapping results to identify splice junctions between exons.

Provide URL(s) to FASTA or FASTQ file(s) with NGS RNA-Seq reads to the input port of the element, set up the reference sequence in the parameters. The result is saved to the specified BAM file, URL to the file is passed to the output port. Several UCSC BED tracks are also produced: junctions, insertions, and deletions.

## Parameters in GUI

Parameter	Description	Default value
<b>Reference input type</b>	Select "Sequence" to input a reference genome as a sequence file. Note that any sequence file format, supported by UGENE, is allowed (FASTA, GenBank, etc.). The index will be generated automatically in this case. Select "Index" to input already generated index files, specific for the tool.	Index
<b>Bowtie index folder</b>	The folder with the Bowtie index for the reference sequence.	
<b>Bowtie index basename</b>	The basename of the Bowtie index for the reference sequence.	
<b>Output folder</b>	The base name of the output folder. It could be modified with a suffix.	
<b>Mate inner distance</b>	The expected (mean) inner distance between mate pairs.	50
<b>Mate standard deviation</b>	The standard deviation for the distribution on inner distances between mate pairs.	20
<b>Library type</b>	Specifies RNA-Seq protocol.	fr-unstranded
<b>No novel junctions</b>	Only look for reads across junctions indicated in the supplied GFF or junctions file. This parameter is ignored if Raw junctions or Known transcript file is not set.	False
<b>Raw junctions</b>	The list of raw junctions.	
<b>Known transcript file</b>	A set of gene model annotations and/or known transcripts.	
<b>Max multihits</b>	Instructs TopHat to allow up to this many alignments to the reference for a given read, and suppresses all alignments for reads with more than this many alignments.	20
<b>Segment length</b>	Each read is cut up into segments, each at least this long. These segments are mapped independently.	25
<b>Fusion search</b>	Turn on fusion mapping.	False
<b>Transcriptome only</b>	Only align the reads to the transcriptome and report only those mappings as genomic mappings.	False
<b>Transcriptome max hits</b>	Maximum number of mappings allowed for a read, when aligned to the transcriptome (any reads found with more than this number of mappings will be discarded).	60
<b>Prefilter multihits</b>	When mapping reads on the transcriptome, some repetitive or low complexity reads that would be discarded in the context of the genome may appear to align to the transcript sequences and thus may end up reported as mapped to those genes only. This option directs TopHat to first align the reads to the whole genome in order to determine and exclude such multi-mapped reads (according to the value of the Max multihits option).	False
<b>Min anchor length</b>	The anchor length. TopHat will report junctions spanned by reads with at least this many bases on each side of the junction. Note that individual spliced alignments may span a junction with fewer than this many bases on one side. However, every junction involved in spliced alignments is supported by at least one read with this many bases on each side.	8
<b>Splice mismatches</b>	The maximum number of mismatches that may appear in the anchor region of a spliced alignment.	0

<b>Read mismatches</b>	Final read alignments having more than these many mismatches are discarded.	2
<b>Segment mismatches</b>	Read segments are mapped independently, allowing up to this many mismatches in each segment alignment.	2
<b>Solexa 1.3 quals</b>	As of the Illumina GA pipeline version 1.3, quality scores are encoded in Phred-scaled base-64. Use this option for FASTQ files from pipeline 1.3 or later.	False
<b>Bowtie version</b>	Specifies which Bowtie version should be used.	Bowtie2
<b>Bowtie -n mode</b>	TopHat uses -v in Bowtie for initial read mapping (the default), but with this option, -n is used instead. Read segments are always mapped using -v option.	Use -v mode
<b>Bowtie tool path</b>	The path to the Bowtie external tool.	default
<b>SAMtools tool path</b>	The path to the SAMtools tool. Note that the tool is available in the UGENE External Tool Package.	default
<b>TopHat tool path</b>	The path to the TopHat external tool in UGENE.	default
<b>Temporary folder</b>	The directory for temporary files.	default
<b>Samples map</b>	The map which divides all input datasets into samples. Every sample has the unique name.	

## Parameters in Workflow File

Type: tophat

Parameter	Parameter in the GUI	Type
reference-input-type	Reference input type	string
bowtie-index-dir	Bowtie index folder	string
bowtie-index-basename	Bowtie index basename	string
out-dir	Output folder	
mate-inner-distance	Mate inner distance	numeric
mate-standard-deviation	Mate standard deviation	numeric
library-type	Library type	numeric
no-novel-junctions	No novel junctions	boolean
raw-junctions	Raw junctions	string
known-transcript	Known transcript file	string
max-multihits	Max multihits	numeric
segment-length	Segment length	numeric
fusion-search	Fusion search	boolean
transcriptome-only	Transcriptome only	boolean
transcriptome-max-hits	Transcriptome max hits	numeric
prefilter-multihits	Prefilter multihits	boolean
min-anchor-length	Min anchor length	numeric
splice-mismatches	Splice mismatches	numeric
read-mismatches	Read mismatches	numeric
segment-mismatches	Segment mismatches	numeric
solexa-1-3-quals	Solexa 1.3 quals	boolean

<b>bowtie-version</b>	<b>Bowtie version</b>	<i>numeric</i>
<b>bowtie-n-mode</b>	<b>Bowtie -n mode</b>	<i>numeric</i>
<b>bowtie-tool-path</b>	<b>Bowtie tool path</b>	<i>string</i>
<b>samtools-tool-path</b>	<b>SAMtools tool path</b>	<i>string</i>
<b>path</b>	<b>TopHat tool path</b>	<i>string</i>
<b>temp-dir</b>	<b>Temporary directory</b>	<i>string</i>

## Input/Output Ports

The element has 1 *input port*.

**Name in GUI:** Input reads

**Name in Workflow File:** in-assembly

**Slots:**

Slot In GUI	Slot in Workflow File	Type
<b>Dataset name</b>	<b>dataset</b>	<i>string</i>
<b>Input reads</b>	<b>first.in</b>	<i>assembly</i>
<b>Input reads url</b>	<b>in-url</b>	<i>string</i>
<b>Input paired reads url</b>	<b>paired-url</b>	<i>string</i>
<b>Input paired reads</b>	<b>second.in</b>	<i>assembly</i>

And 1 *output port*:

**Name in GUI:** TopHat output

**Name in Workflow File:** out-assembly

**Slots:**

Slot In GUI	Slot in Workflow File	Type
<b>Accepted hits</b>	<b>accepted.hits</b>	<i>assembly</i>
<b>Accepted hits url</b>	<b>hits-url</b>	<i>string</i>