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.

Element type: tophat

Parameters

Parameter	Description	Default value	Parameter in Workflow File	Type
Reference input type	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	**reference-input-type**	*string*
Bowtie index folder	The folder with the Bowtie index for the reference sequence.		**bowtie-index-dir**	*string*
Bowtie index basename	The basename of the Bowtie index for the reference sequence.		**bowtie-index-basename**	*string*
Output folder	The base name of the output folder. It could be modified with a suffix.		**out-dir**	
Mate inner distance	The expected (mean) inner distance between mate pairs.	50	**mate-inner-distance**	*numeric*
Mate standard deviation	The standard deviation for the distribution on inner distances between mate pairs.	20	**mate-standard-deviation**	*numeric*
Library type	Specifies RNA-Seq protocol.	fr-unstranded	**library-type**	*numeric*
No novel junctions	Only look for reads across junctions indicated in the supplied GFF or junctions file. This parameter is ignored if R aw junctions or Known transcript file is not set.	False	**no-novel-junctions**	*boolean*
Raw junctions	The list of raw junctions.		**raw-junctions**	*string*
Known transcript file	A set of gene model annotations and/or known transcripts.		**known-transcript**	*string*
Max multihits	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	**max-multihits**	*numeric*
Segment length	Each read is cut up into segments, each at least this long. These segments are mapped independently.	25	**segment-length**	*numeric*
Fusion search	Turn on fusion mapping.	False	**fusion-search**	*boolean*
Transcriptome only	Only align the reads to the transcriptome and report only those mappings as genomic mappings.	False	**transcriptome-only**	*boolean*
Transcriptome max hits	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	**transcriptome-max-hits**	*numeric*
Prefilter multihits	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	**prefilter-multihits**	*boolean*
Min anchor length	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	**min-anchor-length**	*numeric*
Splice mismatches	The maximum number of mismatches that may appear in the anchor region of a spliced alignment.	0	**splice-mismatches**	*numeric*
Read mismatches	Final read alignments having more than these many mismatches are discarded.	2	**read-mismatches**	*numeric*

Segment mismatches	Read segments are mapped independently, allowing up to this many mismatches in each segment alignment.	2	**segment-mismatches**	*numeric*
Solexa 1.3 quals	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	**solexa-1-3-quals**	*boolean*
Bowtie version	Specifies which Bowtie version should be used.	Bowtie2	**bowtie-version**	*numeric*
Bowtie -n mode	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	**bowtie-n-mode**	*numeric*
Bowtie tool path	The path to the Bowtie external tool.	default	**bowtie-tool-path**	*string*
SAMtools tool path	The path to the SAMtools tool. Note that the tool is available in the UGENE External Tool Package.	default	**samtools-tool-path**	*string*
TopHat tool path	The path to the TopHat external tool in UGENE.	default	**path**	*string*
Temporary folder	The directory for temporary files.	default	**temp-dir**	*string*
Samples map	The map which divides all input datasets into samples. Every sample has the unique name.			

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
Dataset name	**dataset**	*string*
Input reads	**first.in**	*assembly*
Input reads url	**in-url**	*string*
Input paired reads url	**paired-url**	*string*
Input paired reads	**second.in**	*assembly*

And 1 *output port*:

Name in GUI: TopHat output

Name in Workflow File: out-assembly

Slots:

Slot In GUI	Slot in Workflow File	Type
Accepted hits	**accepted.hits**	*assembly*
Accepted hits url	**hits-url**	*string*