Find Splice Junction with TopHat Element

TopHat is a fast splice junction mapper for RNA-Seq reads. It aligns RNA-Seq reads to mammalian-sized genomes using the ultra high-throughput short read aligner Bowtie, and then analyzes the mapping results to identify splice junctions between exons.

Parameters in GUI

Parameter	Description	Default value
Output directory	Directory to save MACS output files.	
Bowtie index directory	The directory with the Bowtie index for the reference sequence.	
Bowtie index basename	The basename of the Bowtie index for the reference sequence.	
Mate inner distance	The expected (mean) inner distance between mate pairs.	200
Mate standard deviation	The standard deviation for the distribution on inner distances between mate pairs.	20
Library type	Specifies RNA-Seq protocol.	Standard Illumins
No novel junctions	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
Raw junctions	The list of raw junctions.	
Known transcript file	A set of gene model annotations and/or known transcripts.	
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
Segment length	Each read is cut up into segments, each at least this long. These segments are mapped independently.	25
Fusion search	Turn on fusion mapping.	False
Transcriptome only	Only align the reads to the transcriptome and report only those mappings as genomic mappings.	False
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
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
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
Splice mismatches	The maximum number of mismatches that may appear in the anchor region of a spliced alignment.	0
Read mismatches	Final read alignments having more than these many mismatches are discarded.	2
Segment mismatches	Read segments are mapped independently, allowing up to this many mismatches in each segment alignment.	2
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
Bowtie version	Specifies which Bowtie version should be used.	Bowtie2
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 tool path	The path to the Bowtie external tool.	default
SAMtools tool path	The path to the SAMtools tool. Note that the tool is available in the UGENE External Tool Package.	default
TopHat tool path	The path to the TopHat external tool in UGENE.	default
Temporary directory	The directory for temporary files.	default

Parameters in Workflow File

Type: tophat

Parameter	Parameter in the GUI	Type
out-dir	**Output directory**	*string*
bowtie-index-dir	**Bowtie index directory**	*string*
bowtie-index-basename	**Bowtie index basename**	*string*
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*
bowtie-version	**Bowtie version**	*numeric*
bowtie-n-mode	**Bowtie -n mode**	*numeric*
bowtie-tool-path	**Bowtie tool path**	*string*
samtools-tool-path	**SAMtools tool path**	*string*
path	**TopHat tool path**	*string*
temp-dir	**Temporary directory**	*string*

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*