Assemble Transcripts with StringTie Element

StringTie is a fast and highly efficient assembler of RNA-Seq alignments into potential transcripts. It uses a novel network flow algorithm as well as an optional de novo assembly step to assemble and quantitate full-length transcripts representing multiple splice variants for each gene locus.

Element type: stringtie

Parameters

Parameter	Description	Defaultvalue	Parameter in Workflow File	Type
Reference annotations	Use the reference annotation file (in GTF or GFF3 format) to guide the assembly process (-G). The output will include expressed reference transcripts as well as any novel transcripts that are assembled.		**reference-annotations**	*string*
Reads orientation	Select the NGS libraries type: unstranded, stranded fr-secondstrand (--fr), or stranded fr-firststand (--rf).	Unstranded	**reads-orientation**	*string*
Label	Use the specified string as the prefix for the name of the output transcripts (-l).	STRG	**label**	*string*
Min isoform fraction	Specify the minimum isoform abundance of the predicted transcripts as a fraction of the most abundant transcript assembled at a given locus (-f). Lower abundance transcripts are often artifacts of incompletely spliced precursors of processed transcripts.	0.1	**min-isoform-fraction**	*numeric*
Min assembled transcript length	Specify the minimum length for the predicted transcripts (-m).	200	**min-isoform-fraction**	*numeric*
Min anchor length for junctions	Junctions that don't have spliced reads that align them with at least this amount of bases on both sides are filtered out (-a).	10	**min-anchor-length**	*numeric*
Min junction coverage	There should be at least this many spliced reads that align across a junction (-j). This number can be fractional since some reads align in more than one place. A read that aligns in n places will contribute 1/n to the junction coverage.	1	**min-junction-coverage**	*numeric*
Trim transcripts based on coverage	By default StringTie adjusts the predicted transcript's start and/or stop coordinates based on sudden drops in coverage of the assembled transcript. Set this parameter to "False" to disable the trimming at the ends of the assembled transcripts (-t).	True	**trim-transcripts**	*bool*
Min coverage for assembled transcripts	Specifies the minimum read coverage allowed for the predicted transcripts (-c). A transcript with a lower coverage than this value is not shown in the output. This number can be fractional since some reads align in more than one place. A read that aligns in n places will contribute 1/n to the coverage.	2.5	**min-coverage**	*numeric*
Min locus gap separation	Reads that are mapped closer than this distance are merged together in the same processing bundle (-g).	50 bp	**min-locus-gap**	*numeric*
Fraction covered by multi-hit reads	Specify the maximum fraction of multiple-location-mapped reads that are allowed to be present at a given locus (-M). A read that aligns in n places will contribute 1/n to the coverage.	0.95	**multi-hit-fraction**	*numeric*
Skip assembling for sequences	Ignore all read alignments (and thus do not attempt to perform transcript assembly) on the specified reference sequences (-x). The value can be a single reference sequence name (e.g. "chrM") or a comma-delimited list of sequence names (e.g. "chrM,chrX,chrY"). This can speed up StringTie especially in the case of excluding the mitochondrial genome, whose genes may have very high coverage in some cases, even though they may be of no interest for a particular RNA-Seq analysis. The reference sequence names are case sensitive, they must match identically the names of chromosomes/contigs of the target genome against which the RNA-Seq reads were aligned in the first place.		**skip-sequences**	*string*
Multi-mapping correction	Enables or disables (-u) multi-mapping correction.	Enabled	**multi-mapping-correction**	*bool*
Verbose log	Enable detailed logging, if required (-v). The messages will be written to the UGENE log (enabling of "DETAILS" and "TRACE" logging may be required) and to the dashboard.	False	**verbose-log**	*bool*
Number **of threads**	Specify the number of processing threads (CPUs) to use for transcript assembly (-p).	8	**threads**	*numeric*

Output transcripts file	StringTie's primary output GTF file with assembled transcripts.	Auto	**transcripts-output-url**	*string*
Enable gene abundance output	Select "True" to generate gene abundances output (-A). The output is written to a tab-delimited text file. Also, the file URL is passed to an output slot of the workflow element.	False	**gene-abundance-output**	*bool*

Input/Output Ports

The element has 1 *input port*:

Name in GUI: Input BAM file(s)

Name in Workflow File: in

Slots:

Slot in GUI	Slot in Workflow File	Type
Source URL	url	*string*

And 1 *output port*:

Name in GUI: StringTie output data

Name in Workflow File: out

Slots:

Slot in GUI	Slot in Workflow File	Type
Output URL	url	*string*