Salmon.Indexer Documentation

Description: Indexing a transcriptome in order to perform quantification with Salmon (version 1.5.2) and Alevin. See: The Salmon User Guide for detailed usage guidelines.

Author(s): Rob Patro, COMBINE Lab, Department of Computer Science, University of Maryland. Wrapped as a module by Anthony S. Castanza, Mesirov Lab, UCSD School of Medicine.

Contact: Algorithm and scientific questions: https://github.com/COMBINE-lab/salmon/issues. Module specific issues: https://genepattern.org/help

Summary: In order to quantify gene expression using Salmon or Alevin, the transcriptome must first be indexed. This module performs this indexing step.

Source Publication: Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides fast and bias-aware quantification of transcript expression. Nat Methods. 2017 Apr;14(4):417-419. doi: 10.1038/nmeth.4197. Epub 2017 Mar 6. PMID: 28263959; PMCID: PMC5600148.

Parameters:

Name	Description
GTF gz*	A GTF file containing the genomic features to be indexed. Must be gzipped.
Transcriptome fa gz	A FASTA file of the transcript sequences corresponding to the organism's GTF file. Optional: if a Fasta file is not provided the module will attempt to use gffread to extract the sequences from the Genomic fasta file. Must be gzipped.
Genome fa gz*	A FASTA file of the genomic sequences corresponding to the organism's genome assembly used for the transcript annotation. Must be gzipped.
Kmer*	An odd numbered integer. Used to index the transcriptome, used as the minimum acceptable length of a valid match. (Default: 31, for reads <70bp, consider adjusting this to a smaller value.)
Index Mode*	Options for generating a decoy-aware transcriptome index. Build a full decoy index (full), or a partial decoy index (partial). See parameter description for details.
Use Gencode*	Gencode uses "|" characters in their FASTA record IDs, "true" allows salmon.index to be aware of these characters when processing gene IDs from records
output index name*	The base name for the output indexed transcriptome

*required

Output Files:

Name	Description
<GTF.basename>.k<kmer>.salmon_<Index.Mode>_decoy_index.tar.gz	A gzipped file containing the salmon index for downstream quantification.
<GTF.basename>.fa.gz	If a transcriptome fasta was not provided, this output is generated from the provided GTF file and genome fasta file using GFFread.

Module Language: Shell script

Source Repository: https://github.com/genepattern/Salmon.Indexer/releases/tag/v0.4

Docker image: genepattern/salmon-indexer:beta

Version	Comment
0.4	Initial release.