PreprocessVelocityTranscriptome Documentation

Description: Extract transcript and intron sequences from the genome sequence using the eisaR package in order to quantify both intronic (unspliced) and exonic (spliced) sequences.

Author: Anthony S. Castanza
Contact: genepattern.org/help

Summary: In order to build a transcriptome index for single-cell RNA velocity quantification, intronic (unprocessed) and exonic (processed) RNA sequences must be extracted from the genome. This module prepares the files necessary to produce velocity-compatible input files for the salmon.index module.

Parameters:

Name	Description
GTF	A GTF file containing the genomic ranges to extract features from for quantification.
Genome FASTA	A FASTA file of the genomic sequence corresponding to the organism's GTF file.
Intron Extraction	Consider transcripts separately ("separate") when extracting intronic regions, or collapsed to gene level ("collapse").
Intron Flank Length	Length of sequence flanking introns to allow read alignment across unspliced junctions.

Output Files:

Name	Description
<GTF.basename>.annotation.velocity.<Intron.Flank.Length>bp_flank.gtf.gz	A gzipped GTF file containing the intronic and exonic genomic ranges extracted. Input for the salmon.index module.
<GTF.basename>.annotation.velocity.<Intron.Flank.Length>bp_flank.fa.gz	A gzipped FASTA file of the genomic sequence corresponding to intronic and exonic genomic ranges extracted. Input for the salmon.index module.
<GTF.basename>.annotation.velocity.<Intron.Flank.Length>bp_flank.features.tsv	A two-column file containing the list of spliced gene ids in column 1 and the unspliced gene ids in column 2.
<GTF.basename>.annotation.velocity.<Intron.Flank.Length>bp_flank.tgMap.tsv	A two-column file containing the mappings of transcript level features to gene level features Input for the salmon.alevin.quant module.
<GTF.basename>.annotation.velocity.<Intron.Flank.Length>bp_flank.mtGenes.txt	A list of the gene ids for mitochondrial genes.
<GTF.basename>.annotation.velocity.<Intron.Flank.Length>bp_flank.rrnaGenes.txt	A list of the gene ids with the biotype “rRNA” (ribosomal RNA genes).

Module Language: R 4.0.3
Source Repository: https://github.com/genepattern/PreprocessVelocityTranscriptome/releases/tag/v1
Docker image: jupyter/datascience-notebook:r-4.0.3

Version	Comment
1	Initial release.