Criteria- RNA stability network

Wednesday, October 30, 2019 4:54 PM

Task	Tool
Alignment	HISAT2
Read count	HTSeq-count
Measure RNA Stability	REMBRANDTS

Workflows:

1. Generating read count files:

CRIES (Counting Reads for Intronic and Exonic Segments) compatible with REMBRENDTS
https://github.com/csglab/CRIES

- **Step 1.** Creating GTF annotation files
  REQUIRIES: GTF files
  - For the purpose of inferring stability from RNA-seq, we only consider constitutive exons, i.e. exons that are present in all isoforms of a gene. We also consider only isoforms that are supported by both the Ensembl and Havana annotations, as provided in the GTF file.

- **Step 2.** Mapping reads
  REQUIRIES: fastq files
  - We use HISAT2 for mapping RNA-seq reads, keeping only alignments with MAPQ score ≥30

- **Step 3.** Counting reads that map to intronic or exonic segments of each gene
  REQUIRIES: bam files
  - We use **HTSeq-count** for counting reads.
  - For counting exonic reads,
    - we run the HTSeq-count using the "intersection-strict" mode, to ensure that the reads that are counted entirely fit within the mature mRNA sequence, and do not overlap alternatively spliced exons.
  - For intronic reads,
    - we run the HTSeq-count using the "union" mode, since any read that even partially overlaps an intronic region likely originates from pre-mature RNA.

- **Step 4.** Normalization
  REQUIRIES: count tables
  - In order to obtain stability measurements we can use REMBRANDTS package for
    - read count normalization,
    - gene filtering,
    - removing bias
  - REMBRANDTS internally uses DESeq, and performs linear regression of
    $$\Delta_{exon} - \Delta_{intron} \text{ vs. } \Delta_{intron}$$
    to obtain residuals that correspond to unbiased estimates of stability.

=> In the read count files, each row corresponds to one gene, with the

- First column representing the gene ID
- The second column representing the total number of reads mapped to that gene (either

intronic or exonic reads).

2. Generate cystoscope network