Strategies for Bulk RNA-seq Analysis
Genome Mapping

Reads

STAR, HISAT2

Splice-aware Genome mapping

htseq-count, featureCounts

Gene counting

Transcript discovery & counting

Homology-based BLAST2GO

Novel transcript annotation

Transcriptome Mapping

Reads

RSEM, Kallisto, Sailfish, Salmon

Transcript mapping and quantification

StringTie

Transcript counting

Assembly

Reads

Trinity, Scripture, Stringtie

Assembly into transcripts

Trinotate

Novel transcript annotation
✓ Genome
✓ GTF (annotation)

Sequence reads

FASTQ

Quality control

FASTQ (+reference genome index) (+known GTF, optional)

Alignment to Genome: HISAT2, STAR

multiple BAMs (+known GTF)

Count reads associated with genes: htseq-count, featureCounts

Count Matrix

DGE with R: DESeq2, EdgeR, limma:voom
✓ Genome
✓ GTF (annotation)?

Sequence reads

Quality control

FASTQ

(+known GTF, optional)

Alignment to Genome: HISAT2, STAR

multiple BAMs

Count reads associated with genes: htseq-count, featureCounts

Count Matrix

DGE with R: DESeq2, EdgeR, limma:voom

DGE with CuffDiff, Ballgown

Reference-based transcriptome assembly and quantitation with StringTie

BAM

FASTQ

(+reference genome index)
✓ Transcriptome (FASTA)

Sequence reads
- FASTQ
- Quality control
- FASTQ
- Alignment to Genome: HISAT2, STAR
- multiple BAMs (+known GTF)
- Count reads associated with genes: htseq-count, featureCounts
- Count Matrix

Reference-based transcriptome assembly and quantitation with StringTie

DGE with CuFF, Ballgown

Pseudocounts with Kallisto, Sailfish, Salmon

DGE with R: DESeq2, EdgeR, limma:voom

Count Matrix generated using tximport

DGE with Sleuth

FASTQ (+reference transcriptome index)
Reference-based assembly

✓ Genome
✓ GTF (annotation)?

De novo assembly

✓ Genome?
✓ GTF (annotation)?

Quantitation from assembled reads

Alignment to new transcriptome: Bowtie2, BWA

SAM/BAM

Count reads associated with genes

Count Matrix

DGE with R: DESeq2, EdgeR, limma:voom

Novel transcript annotation

Trinotate

Assembly into transcripts

Trinity, Scripture

RSEM, Kallisto, Salmon, eXpress

Transcript mapping & quantification

Reads

Counts reads associated with genes
These materials have been developed by members of the teaching team at the Harvard Chan Bioinformatics Core (HBC). These are open access materials distributed under the terms of the Creative Commons Attribution license (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.