omics.data.edu.au

Transcriptomics Requirements Summary

This document is a high level summary of the Transcriptomics requirements and information captured during the discoverability phase of the RDS Omics project

Summary

Transcriptomics is the study of the transcriptome—the complete set of RNA transcripts that are produced by the genome, under specific circumstances or in a specific cell—using technologies such as high-throughput sequencing. (www.nature.com/subjects/**transcriptomics**)

Groups Consulted

- AGRF c/o WEHI, 1G Royal Parade Parkville VIC 3050 (Sonika Tyagi)
- VLSCI, LAB-14, 700 Swanston St Carlton VIC 3052 (Torsten Seemann)

Work Pattern

1. Obtain sequence reads
2. Choose annotated reference genome
3. For each sample's replicate
    a. Align reads to reference
    b. Count reads mapped to each gene feature
4. Combine counts into single file against each gene
5. Perform statistical analysis
6. Obtain a list of differentially expressed genes
7. Visualize results

Data Overview

- Each strain's RNA will be sequenced in 2 conditions with 5 replicates each, giving 10 "samples"
- Each sample consists of two FASTQ files (paired-end sequencing)

Software/Applications used

- Read alignment: *BWA MEM, Bowtie, Kallisto*
- Read counting: *htseq-count, Kallisto*
- Statistical analysis: *Voom/Limma, EdgeR, DESeq2*
- Visualization: *Degust, JBrowse*

Deficiencies in processes and tools

Primary work pattern from FASTQ to VIsualization is well understood and follows best practice. Further analysis such as pathway analysis, gene set enrichment tests, and linking to proteomic data is more difficult and outside the scope of this project.

Relationship with other Streams

- Requires a finished genome and good annotation from the genomics stream.
- Alignments to unannotated regions can be used to curate/correct/update the annotation.
- Differential expression results can be used for proteomics and metabolomics analysis.