**The RNA-Seq Analysis Workflow**



**Overview**

The current version of this walkthrough was developed with help of Colin Dewey and provides an overview of the analysis pipeline from pre-processed fastq files (Raw sequence data and QC report provided by Novogene) to the generation of gene and isoform matrices for further downstream analyses.

To meet the high computational demand of the analyses process, the first two steps of the pipeline (STAR and RSEM) have been modified to run on the UW Madison’s Centre for High Throughput Computing (CHTC) platform, Condor.

I have provided in addition to the main scripts, the actual job submission files used. For both steps, I utilize the docker containers with up-to-date image versions of the software, to run jobs and skip the process of software packaging and shipping to CHTC’s condor environments.

**Read Alignment Mapping**

In the STAR folder I have included all working scripts for the genome alignment. The **STAR\_indexing.sh**, which generates index files needed for the alignment and the **STAR\_align.sh** which performs the genome alignment. The **.sub** files are the job submission files used for executing both jobs.

**Quantification of Gene/Transcript Expression**

Similarly, the RSEM folder includes scripts used for running RSEM. **RSEM\_indexing.sh** generates index files for RSEM, which **RSEM\_exec.sh** utilizes for quantification.

**Differential Gene Expression Analysis**:

Gene/isoform quantification and identification using the DESeq2 R package. I have also included the **Diff\_Expr\_Analyses.R** script which generates vst-normalized matrices of genes and isoforms.