

# RNA-seq Standard Analysis Workflow & Results Delivery & Timeline Policy

## Purpose

To establish a clear and structured standard analysis timeline for delivering analysis results from RNA-seq experiments, ensuring timely communication of progress and findings to all clients.

## Scope

This policy applies to all standard workflow RNA-seq analysis projects undertaken by the Bioinformatics Core, Center for Research Informatics, Biological Science Division, the University of Chicago.

## Workflow Stages & Tools Selection

Our standard analysis workflow is listed below with 3 stages as shown in the attached workflow page. This workflow is our **default standard** analysis workflow with estimated **10-15 hours** for the analysis based on your data complexity, **alternative tools can be chosen and added upon the request by all clients in appendix with extra analysis hours.**

### 1. Data Download or Transfer to HPC

- sftp/lftp
- Globus

### 2. Nextflow Pipeline Analysis (HPC computation), Steps Included:

- ◆ Initial Quality Control
  - FastQC: Raw sequencing QC
  - trimGalore: Adapter & low-quality reads trimming
  - BBSplit: Genome contamination removal
  - SortMeRNA: Ribosomal RNA removal
- ◆ Read Alignment
  - STAR
- ◆ Post-alignment quality inspection
  - RSeQC: Post-alignment quality inspection
  - Picard: Post-alignment quality inspection
- ◆ Genome Assembly

- salmon
- 3. Downstream Analysis (Local PC), Steps Included:**
  - ◆ Differential expression analysis
    - DEseq2
  - ◆ Gene set enrichment and pathway analysis
    - ClusterProfiler

## Delivery Milestones

### 1. Timeline

- **Data Download or Transfer to HPC:** Within 1-2 business days of initiating raw sequencing data transfer request by the client.
- **Nextflow Pipeline Analysis (HPC computation):** Within 1 week after data successfully downloaded to HPC (randi)
- **Downstream Analysis (Local PC):** Within 1 week after execution of next-flow pipeline

### 2. Deliverables

- **Data Download or Transfer to HPC**
  - Raw sequencing data (\*.fastq.gz files) transfer (**Optional with extra hours**)
- **Nextflow Pipeline Analysis (HPC computation)**
  - Alignment files (\*.bam), (**Optional with extra hours**)
- **Downstream Analysis Html Report (Local PC)**
  - Gene expression count table (\*.featurecount.txt/.csv/.excel)
  - Differentially expressed genes for all comparisons, including
    - Summary table of number of DEGs at  $FDR \leq 0.05$  and  $|FC| \leq 1.5$
    - DEGs table (\*.comparison\*.txt/csv/excel)
    - Volcano plot (\*.comparison\*.pdf)
  - Enriched functional analysis results
    - Full clusterProfiler analysis results table (\*.clusterProfiler.txt/csv/excel)
    - Dot-plot of enriched functions/pathways (\*.pdf)

## Reporting and Communication

### 1. Interim Communication

- Interim reports will be provided at the completion of each major workflow stage, summarizing key findings and any issues encountered for review and discussion.
- Interim reports for review will be within 1 business day following the completion of each respective stage.

## **2. Final Report**

- All deliverable results will be shared via box sharing by default.
- The final report will be compiled from deliverable results of all stages of the workflow presented above in html markdown files.
- Delivery of the final report will be within 2 business days after the completion of the final analysis stage.

## **Follow-up Analysis**

Any further follow-up requests regarding publication, paper revision and grant preparation can be conducted and will be charged by additional hours.

## **Compliance and Quality Assurance**

All analysis steps must comply with the standard operating procedures and quality assurance guidelines established by our core. Any deviations or delays from the timeline must be documented and communicated to the project stakeholders with justifications and revised delivery dates.

## **Policy Review**

This policy will be reviewed annually to ensure it remains relevant and up-to-date with the latest technological advancements and organizational requirements.

# Standard Analysis Workflow

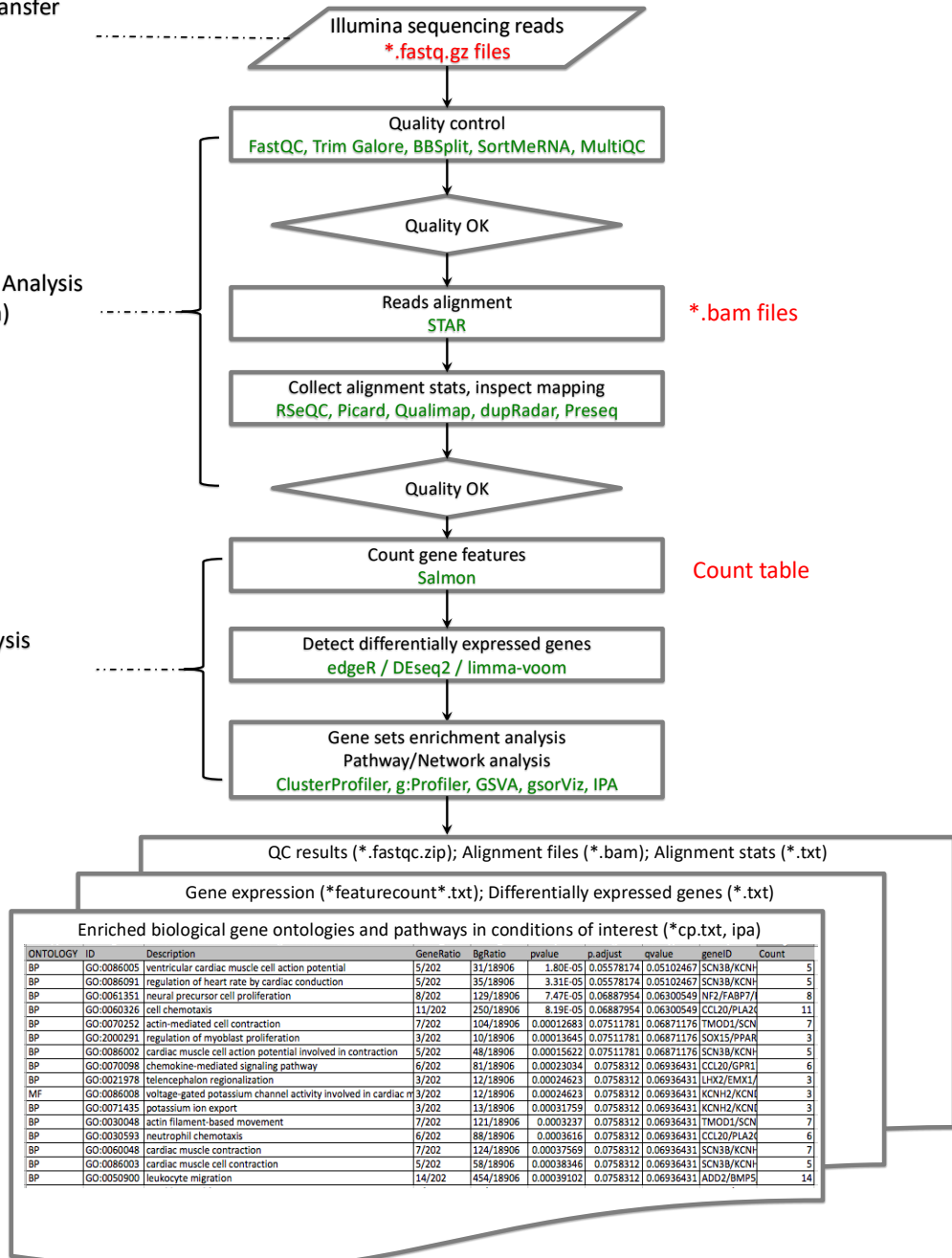


## CRI RNA-seq Analysis Pipeline

Data Download/Transfer  
(sftp, lftp, golbus)  
1-2 days

Nextflow Pipeline Analysis  
(HPC computation)  
1 week

Downstream Analysis  
(Local PC)  
1 week



## **Appendix**

### **1. Data/Results Download or Transfer**

- Globus
- Hard drive

### **2. Nextflow Pipeline Analysis (HPC computation), Steps Included:**

- ◆ Read Alignment
  - BWA
  - Bowtie
- ◆ Post-alignment quality inspection
  - Picard tools
  - samtools
  - UCSC genome browser
  - IGA
- ◆ Genome Assembly
  - featureCounts
  - cufflinks
  - Kallisto

### **3. Downstream Analysis (Local PC), Steps Included:**

- ◆ Differential expression analysis
  - edgeR
  - limma-voom
  - ComBat (batch effect removal)
- ◆ Functional enrichment analysis
  - GSEA
  - g:Profiler
  - GSEA

### **4. Sequencing Data Submission**

- GEO
- EBI

### **5. Customized Requests:**

---

---

---

---

---

---

---