RNA-Rocket

The RNA-Seq Pipeline at

Pathogen Portal
The Bioinformatics Resource Centers Portal

CyberInfrastructure Division
Thank you.
Pathogen Portal

Featuring...

**RNA Rocket**
Align your Illumina fastQ reads against supported genomes, view supported genomes, and estimate gene expression values using an RNA-Seq Pipeline running on Galaxy.

- New user interface
- New reference genomes
- New tools

**Pathogen Interaction Gateway**
Generate a network graph of Protein-Protein Interactions, including Host-Pathogen Interactions, from your custom selection of hosts/vectors, bacteria, viruses, and eukaryotic pathogens.

Explore Infectious Disease
RNA-Rocket
RNA–Rocket

- Free storage 300GB
- Nightly updates from VectorBase, EuPathDB, PATRIC
- 6824 bacteria
- 60 Eukaryotic pathogens
- 8 model organisms
- 6 vectors
RNA-Seq Pipeline: Initial Release January 2012

Analyse
Explore and use resources for analyzing host response to infectious disease.

RNA-Seq Pipeline
Map your RNA-Seq Reads to Reference Genomes
- Align your Illumina fastq reads (pipped fastq files accepted) against any sequenced genome from
  RefSeqDB, PATRIC, and VectorBase.
- View a list of supported genomes.
- Estimate Gene Expression Values.
- Obtain BAM files for the resulting alignments and FPKM expression values for annotated genes and novel transcripts.

Mouse Model Selection Guide
The Mouse Model Strain Selection Guide was developed by the Pathogen Portal Team in collaboration with The Jackson Laboratory. The Guide lists pathogens along with mouse strains that have been found to be either susceptible or resistant to infection with the pathogen. Links are provided to more information about the mouse strains, including ordering information. Links are also provided to publications documenting susceptibility or resistance of the mouse strains to specific pathogens.
RNA-Seq Pipeline Features

- Guided tour for naïve users

Choose an activity below

**Uploads**
- Upload files
  - Upload files for analysis via URL, FTP, or HTTP.

**Quality Control**
- Check read quality
  - Optional: Run FastQC to get a report on the quality of base calls that could affect your read mapping.
  - Trim Reads
  - Optional: Run the sickle trimming tool to trim your reads and prepare them for alignment.
  - Check mapping quality
  - Optional: Check the number of reads mapped and the alignment quality.

**RNA-Seq Analysis**
- Align Reads & Assemble Transcripts
  - Map your reads to the genome and assemble them into transcripts. The alignment step will generate BAM files and the assembly step will generate BED and GTF files.
- Differential Expression Analysis
  - Test RNA-Seq samples to determine if transcripts are differentially expressed.
- Create GeneList Summary
  - Create a GenList file for use in PATRIC and other differential expression analysis tools.

**Additional Tools**
- BED Tools
  - Use BEDTools to create summary BED files for analysis of genome coverage.
- Remove Duplicate Reads
  - Optional: PCR amplification can lead to bias. For paired-end reads only, if multiple pairs of reads have the exact same coordinates mark all except one as a duplicate and remove.
- Alignment Only
  - To use advanced alignment parameters and/or perform alignment against a non-BRC genome.
Features

- Guided tour for naïve users
RNA-Seq Pipeline Features

Guided tour for naïve users

Choose an activity below

Uploads

Upload Files
Upload files for analysis via URL, FTP, or HTTP.

Quality Control

Check read quality
Optional: Run FastQC to get a report on the quality of base-calls that could affect your read mapping.

Trim Reads
Optional: Run theickle trimming tool to trim your reads and prepare them for alignment.

Check mapping quality
Optional: Check the number of reads mapped and the alignment quality.

RNA-Seq Analysis

Align Reads & Assemble Transcripts
Map your reads to the genome and assemble them into transcripts. The alignment step will generate BAM files and the assembly step will generate BED and GFF files.

Differential Expression Analysis
Test RNA-Seq samples to determine if transcripts are differentially expressed.

Create GeneList Summary
Create a GeneList file, for use in PATRIC and other differential expression analysis tools.

Additional Tools

BED Tools
Use BEDTools to create summary BED files for analysis of genome coverage.

Remove Duplicate Reads
Optional: PCR amplification can lead to bias. For paired-end reads only: if multiple pairs of reads have the exact same coordinates mark all except one as a duplicate and remove.

Alignment Only
To use advanced alignment parameters and/or perform alignment against a non-BRC genome.
RNA-Seq Pipeline Features

- Guided tour for naïve users

Align Reads & Assemble Transcripts

Purpose:
This procedure will map RNA-Seq reads to one of the provided reference genomes and use this mapping to assemble transcripts, map transcripts to existing annotations, and determine the level of expression. Choose the appropriate options for your organism (Eukaryotic/Prokaryotic) and read type (Paired-end/Single-end) below.

Required Input:
FastQ files

Output:
Read alignments (BAM files), tab-delimited assembly and expression files for known genes, isoforms, and novel transcripts.

Select Analysis Type
- Eukaryotic Single-End Analysis
- Eukaryotic Paired-End Analysis
- Prokaryotic Single-End Analysis
- Prokaryotic Paired-End Analysis

Select an existing Project or create a new Project to be used during this analysis and populate the Project with the necessary files. Output from this analysis will be saved in the selected Project.

Currently Selected Project: None Selected

Target Project:
- Select existing project
- CREATE PROJECT

Source Project:
Select source

- Upstream_READ1.fq.gz
- Downstream_READ2.fq.gz
- Trimmed_Upstream_READ1.fq.gz
- Trimmed_Downstream_READ2.fq.gz
- BaseQuality_Upstream_READ1.fq.gz.html
- BaseQuality_Downstream_READ2.fq.gz.html
- Align with Bowtie2 Original
- Align with Bowtie2 Trimmed
- Samtools for Align with Bowtie2 on Original
- Samtools for Align with Bowtie2 Trimmed.html
RNA-Seq Pipeline Features

Quality control tools for read data

Statistics for Align with Bowtie2 on data 1 and data 2: mapped reads

Mapping stats: 73% aligned (25.3M aligned out of 34.5M total)

Mean Base Quality

Quality scores across all bases (Illumina 1.5 encoding)
RNA-Seq Pipeline Features

- BRC streaming and viewing of results
Galaxy Modifications

- Launch Pad
  - D3 library concept diagram
Galaxy Modifications

- Launch Pad
  - launch_pad.py controller
  - Tool and workflow lists
Galaxy Modifications

- Launch Pad
  - launch_pad.py controller
  - Tool and workflow lists

<a href="${h.url_for(controller='launch_pad', action='leave_hanger', launch_type='workflow', retrieval_tag='fastqc')}">Check read quality</a>
Launch configuration
- Tool/Workflow lists
- Gateway user
- API copy
Galaxy Modifications

Configure Workflow Run for "CSU Demo"

- Workflow configuration
  - Paired-end workflows
  - Extjs- run.mako
  - Coordinated selection
    - Display type
  - AJAX load
  - Cufflinks parameters e.g.
Galaxy Modifications

- Workflow configuration
  - Paired-end workflows
  - Extjs– run.mako
  - Coordinated selection
    - Display type
  - AJAX load
  - Cufflinks parameters e.g.
Other Modifications

- Cuffdiff/cuffmerge integration
- SAM Stat
- Select2 wait time
- Power users
Challenges

- Controller system
- Workflow vs tool runners
  - Fixed parameters
  - Inputs
- Advanced parameters
Future Plans

- Visualization of results
  - CummeRbund
- Organism specific tools
- Link outs to BRCs
  - Annotation differences
  - Differential Expression Visualization
  - Alignment results streaming
Acknowledgements

c pathogenportal.org

- Eric Nordberg
- Dustin Machi
- Chunhong Mao

- This project has been funded in whole or in part with Federal funds from the National Institute of Allergy and Infectious Diseases, National Institutes of Health, Department of Health and Human Services, under Contract No. HHSN272200900040C, awarded to BWS Sobral.