Approaches for small RNA-seq in Galaxy

30 June
GCC2017

Mallory Freeberg
Johns Hopkins University
Analysis of small RNAs in Galaxy

- Background
- Experiment
- Approach
- Galaxy tutorial/workflows
- Results
- Continued work
What is a small RNA?

- 18-40 nt ssRNA
- Complexes with Argonaute family proteins
- Post-transcriptional regulation
  - Also transcriptional regulation...
- Target RNA silencing
  - Also protecting...
- Transcribed from DNA and RNA
- Distinct from other (sightly longer) noncoding RNAs
  - rRNA, tRNA, snoRNA, etc.
Small RNAs are diverse

DOI: 10.1016/j.bbagrm.2013.11.009
Why is small RNA analysis difficult?

• Small RNA diversity
  – Characteristics/organism
  – Function
  – Desired results
• Ad hoc downstream analyses
• Study-specific hypotheses

**Difficulty != Exclusion**

Generate GTN tutorial for small RNA-seq analysis
Analysis of small RNAs in Galaxy

• Background
• **Experiment**
• Approach
• Galaxy tutorial/workflows
• Results
• Continued work
Training data acquired from published study

- Small RNA-seq experiments
  - Size selection and rRNA depletion
  - Triplicates (yay!)
  - Control and knockdown of potential small RNA processing component in *Drosophila*
- mRNA-seq experiments in parallel (yay!)
- Questions
  - How are small RNAs levels affected when component is knocked down?
  - Are corresponding targets differentially abundant?
Key considerations for small RNA-seq analyses

• Adaptor trimming: Yes
  Quality trimming: No
• Custom annotation references
  – Transposons, small RNA clusters, transcripts
• Hierarchical alignment to and removal of artifacts
  – e.g. ribosomal RNAs
• Quantification of sense and antisense alignments separately
• Parallel RNA-seq experiments are necessary to assess functional effects of small RNAs
Analysis of small RNAs in Galaxy

- Background
- Experiment
- **Workflow (general)**
- Galaxy tutorial/workflows
- Results
- Continued work
Generalized small RNA-seq workflow

- Control:
  - Input FASTQ files (demultiplexed)

- Condition:
  - Input FASTQ files (demultiplexed)

1. Convert FASTQ to Sanger/Illumina 1.9, if needed
2. Trim off adaptor sequences
3. Remove rRNA reads
4. Separate non-miRNA reads into subclasses based on feature (e.g., length, 5’ nt)
5. Quantify sense and antisense reads against reference features (e.g., transcripts, TEs)
6. Test for differential abundance
7. Assess, visualize results, other downstream analyses
8. Identify miRNA reads and remove for separate analyses
9. Quantify pre/pri/mature miRNA expression
10. Predict targets, other downstream analyses
Available small RNA-seq analysis pipelines

- Many are command-line
- None incorporate RNA-seq

**piPipes**

**shortran**

**UEA small RNA Workbench**
Analysis of small RNAs in Galaxy

- Background
- Experiment
- Workflow
- Galaxy tutorial/workflows available:
  
  https://galaxyproject.github.io/training-material/topics/transcriptomics/tutorials/srna/

- Results
- Continued work
Galaxy Training!

Collection of tutorials developed and maintained by the worldwide Galaxy community.

Galaxy Training Network helps researchers find online training material, and trainers in their geographic area, and helps advertise training events as well.

Transcriptomics

Training material for all kinds of transcriptomics analysis.

Requirements

Before diving into this topic, we recommend you to have a look at:

- Galaxy introduction
- Quality control

Material

<table>
<thead>
<tr>
<th>Lesson</th>
<th>Hands-on</th>
<th>Slides</th>
<th>Input dataset</th>
<th>Galaxy tour</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction</td>
<td></td>
<td><img src="slides.png" alt="Image" /></td>
<td></td>
<td><img src="galaxy-tour.png" alt="Image" /></td>
</tr>
<tr>
<td>De novo transcriptome reconstruction with RNA-seq</td>
<td><img src="slides.png" alt="Image" /></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reference-based RNA-seq data analysis</td>
<td><img src="slides.png" alt="Image" /></td>
<td><img src="input-dataset.png" alt="Image" /></td>
<td></td>
<td><img src="galaxy-tour.png" alt="Image" /></td>
</tr>
<tr>
<td>Differential abundance testing of small RNAs</td>
<td><img src="slides.png" alt="Image" /></td>
<td><img src="input-dataset.png" alt="Image" /></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
sRNA-seq Galaxy workflow Part 1: Preprocessing and removal of artifacts
sRNA-seq Galaxy workflow Part 2: Transcript quantification and DA testing
Analysis of small RNAs in Galaxy

- Background
- Experiment
- Workflow
- Galaxy tutorial/workflows
- Results (DA testing)
- Continued work
DA of antisense siRNAs detected
Sig. increased abundance of antisense siRNAs against **TEs** (+ **Symplekin**)

<table>
<thead>
<tr>
<th>GenelID</th>
<th>Base mean</th>
<th>log2(FC)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NM_141323.2</td>
<td>386.937430190785</td>
<td>7.30054014682614</td>
</tr>
<tr>
<td>NR_133553.1</td>
<td>351.498478664811</td>
<td>-1.48480458529902</td>
</tr>
<tr>
<td>NM_001276174.1</td>
<td>28.9598220633042</td>
<td>2.71241476031759</td>
</tr>
<tr>
<td>FBgn0000004_17.6</td>
<td>253.485313867193</td>
<td>0.931644156005955</td>
</tr>
<tr>
<td>FBgn0014947_flea</td>
<td>105.554020985323</td>
<td>1.26669739564706</td>
</tr>
<tr>
<td>FBgn002697_mdg1</td>
<td>371.126947585139</td>
<td>0.761999311890377</td>
</tr>
<tr>
<td>FBgn002698_mdg3</td>
<td>243.34404785899</td>
<td>0.775788049657332</td>
</tr>
<tr>
<td>NM_001297944.1</td>
<td>262.240211839715</td>
<td>-0.703885941483931</td>
</tr>
<tr>
<td>NR_001639.1</td>
<td>293.573176779384</td>
<td>-0.646840711127527</td>
</tr>
<tr>
<td>FBgn0004904_TART-B</td>
<td>83.2234143548566</td>
<td>1.05935646352502</td>
</tr>
<tr>
<td>NM_143464.4</td>
<td>8.64898787929223</td>
<td>2.27313194212603</td>
</tr>
<tr>
<td>FBgn0001283_jockey</td>
<td>363.917863170095</td>
<td>0.527113612698375</td>
</tr>
<tr>
<td>FBgn0004082_Tirant</td>
<td>152.771195074136</td>
<td>0.720031072809832</td>
</tr>
<tr>
<td>FBgn0004904_TART-C</td>
<td>20.1516466966232</td>
<td>1.63650624789203</td>
</tr>
<tr>
<td>NM_170478.3</td>
<td>6.65127460209622</td>
<td>2.09490499790299</td>
</tr>
</tbody>
</table>
Analysis of small RNAs in Galaxy

- Background
- Experiment
- Workflow
- Galaxy tutorial/workflows
- Results
- Continued transcriptomics work in Galaxy
Improving transcriptome tutorials in GTN

- Parallel mRNA-seq analysis using Salmon
- Integrative tutorials
  - small RNA-seq + RNA-seq
  - ChIP-seq + RNA-seq
- Writing/wrapping additional small RNA-seq tools
- Additional txomics-related content in GTN
  - CLIP-seq
  - Visualization tools, IE
  - Nanopore sequencing
Acknowledgements

GCC2017 organizing committee
Galaxy Training Network

Amazon Web Services
XSEDE/Jetstream
The Galaxy Team
James Taylor lab

Avez-vous des questions?
Merci de votre attention!
Small RNA-seq replicates cluster well
Small RNA-seq replicates cluster well
Few RNA features exhibit “significant” DA