



Banana Slug Analytics Platform

Exosome Small RNA-Seq Analysis Results

This document provides a guide to explore the Exosome Small RNA-Seq analysis results on the Banana Slug Analytics Platform. This should be expected to be used after the analysis has been performed.

At the HOME page of your account, select the project in the Projects panel on the left that has the analysis performed. Then click on the analysis listed under the Analyses panel on the right.

The ANALYSIS page will be opened with all the results displaced at each corresponding stage of the analysis workflow: Analysis Configuration, Quality Assessment, Read Alignment, Abundance Determination, Differential Expression. You can expand the desired stage on the left to explore the data.

Each stage has three types of presented information:

- Status: Current status of the analysis stage with start and completion timestamps
- Output: Results produced at the corresponding analysis stage
- Details: Detailed information of analysis stage, for example, major software command and options used during the analysis

Analysis Configuration

Users can download the sample configuration file that includes the sample/replicate names and their corresponding FASTQ files used in the analysis. This is especially useful for analysis that includes multiple replicates for each treatment or tissue sample.

- ✔ Analysis Configuration ▼
 - [Status](#) • [Output](#) • [Details](#)
- ✔ Quality Assessment ◀
- ✔ Read Alignment ◀
- ✔ Abundance Determination ◀
- ✔ Differential Expression ◀

October 17th 2018, 9:20:26 am
Sample replicate configuration

[Download Sample Configuration](#)

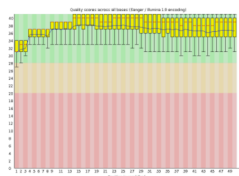
Quality Assessment

This stage includes checking the quality of the raw sequencing data, trimming low quality reads, and computing statistics of read trimming results. Users can enlarge plots for visualization by clicking on the plots or designated links or download plots on the FILES page of the platform.

- ✔ Analysis Configuration ◀
- ✔ Quality Assessment ▼
 - [Status](#) • [Output](#) • [Details](#)

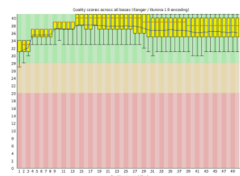
Sequencing data is analyzed for quality and contamination. Raw reads are trimmed to remove adapters, then trimmed and filtered based on quality scores. The scores used for trimming and filtering are specific to the sequencing platform. The preprocessed reads are then assessed for quality and plots are generated of per base quality before and after trimming.
- ✔ Read Alignment ◀
- ✔ Abundance Determination ◀
- ✔ Differential Expression ◀

October 17th 2018, 10:51:10 am
Quality Assessment Before and After Filtering and Trimming for Day8_replicate2



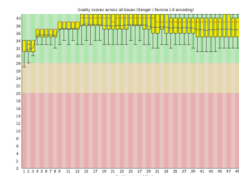
Distribution of the quality scores for each position. The interquartile range (yellow box) and the mean and median values

October 17th 2018, 10:51:10 am
Quality Assessment Before and After Filtering and Trimming for Day14_replicate2



Distribution of the quality scores for each position. The interquartile range (yellow box) and the mean and median values

October 17th 2018, 10:51:24 am
Quality Assessment Before and After Filtering and Trimming for Day2_replicate1



Distribution of the quality scores for each position. The interquartile range (yellow box) and the mean and median values

Read Alignment

This stage includes aligning trimmed reads from the quality assessment stage to the reference genome and computing read mapping statistics. The outputs provide the number of reads used for mapping, number of alignments, number of aligned reads, and percent of aligned reads in each sample/replicate. Summary bar charts are included to show the mapping rates of all the samples in the analysis.

✓ Analysis Configuration

✓ Quality Assessment

✓ Read Alignment

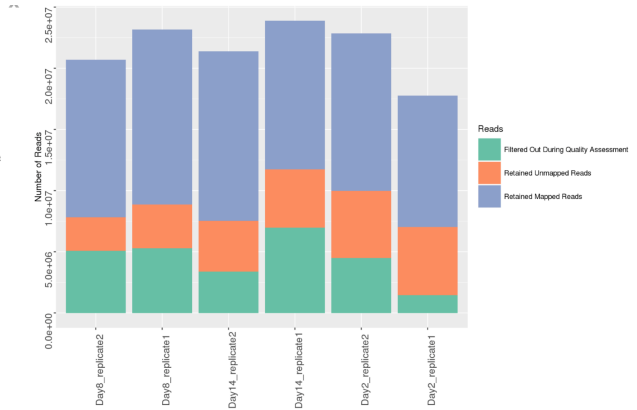
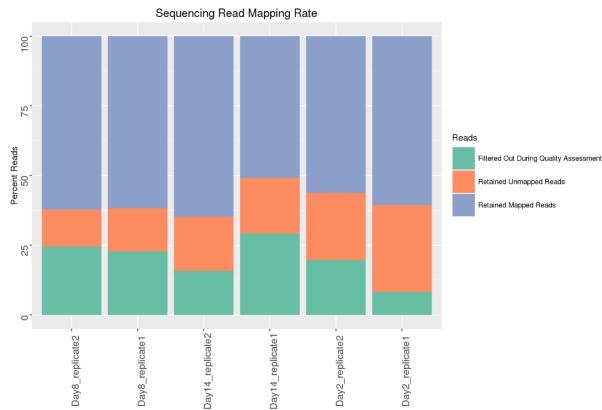
Status • Output • Details

Sequence reads are mapped to the genome and mapping statistics are generated. BAM files with chromosomal location of mapped reads and the alignment quality are made available for visualization in the integrated UCSC Genome Browser.

✓ Abundance Determination

✓ Differential Expression

<p>October 17th 2018, 1:14:31 pm BAM Alignment for Sample Day14_replicate2</p> <p>Download BAM File</p> <p>October 17th 2018, 1:19:19 pm BAM Alignment for Sample Day2_replicate2</p> <p>Download BAM File</p>	<p>October 17th 2018, 1:15:05 pm BAM Alignment for Sample Day8_replicate2</p> <p>Download BAM File</p> <p>October 17th 2018, 1:28:09 pm Mapping Summary for Day14_replicate2:</p> <p>Number of reads used for mapping: 18018810</p> <p>Number of alignments: 83050491</p> <p>Number of aligned reads: 13855365</p> <p>Percent of reads aligned: 76.89</p>	<p>October 17th 2018, 1:19:16 pm BAM Alignment for Sample Day8_replicate1</p> <p>Download BAM File</p> <p>October 17th 2018, 1:28:10 pm Mapping Summary for Day8_replicate2:</p> <p>Number of reads used for mapping: 15632914</p> <p>Number of alignments: 93202781</p> <p>Number of aligned reads: 12878243</p> <p>Percent of reads aligned: 82.37</p>	<p>October 17th 2018, 1:28:54 pm Mapping Summary for Day8_replicate1:</p> <p>Number of reads used for mapping: 17888970</p> <p>Download BAM File</p>
--	---	--	---



Abundance Determination

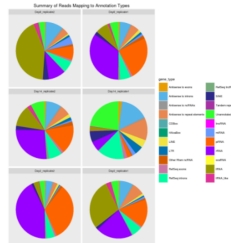
This stage includes estimating the abundance of each annotated gene feature in the reference genome for each sample/replicate. Users can obtain the summary information through the pie charts or download the raw read counts.

- ✔ Analysis Configuration ◀
- ✔ Quality Assessment ◀
- ✔ Read Alignment ◀
- ✔ Abundance Determination ▾

[Status](#) • [Output](#) • [Details](#)

Abundance levels for ncRNAs (miRNAs, tRNAs, rRNAs, lincRNAs, piRNAs, snoRNAs), antisense transcripts, coding genes and repeat elements (LTR, LINE, SINE, and tandem repeats) are determined. A summary of reads overlapping each of these annotations is created and provided in the form of pie-charts.

October 17th 2018, 2:18:32 pm



Pie-chart summarizing the number of reads mapping to each annotation type in each sample

October 17th 2018, 2:18:45 pm

Raw Read Counts

[Download Raw Read Counts File](#)

Differential Expression

This stage includes the differential expression analysis results using DESeq. Users can download the DESeq output files, visualize the pairwise analysis volcano plots of fold-changes and adjusted P-values of all the genes, and launch the interactive DESeq heatmap with the UCSC Genome Browser.

- ✔ Analysis Configuration ◀
- ✔ Quality Assessment ◀
- ✔ Read Alignment ◀
- ✔ Abundance Determination ◀
- ✔ Differential Expression ▾

[Status](#) • [Output](#) • [Details](#)

Differences in expression of ncRNA, antisense transcripts, and repeat elements between samples are calculated. Visual representation of the analysis results are provided, including interactive tabular and heat map views linked to the integrated genome browser.

October 17th 2018, 2:36:14 pm
DESeq output

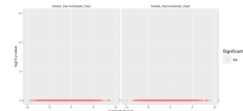
[Download DESeq Output Files](#)

October 17th 2018, 2:37:20 pm
Volcano Plots showing fold-change and p-values of all genes



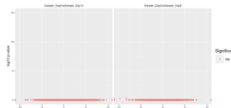
Fold-change (log2) versus significance (-log10 p-value) for each gene is shown. Significant genes (FDR < 0.05) are in blue.

October 17th 2018, 2:37:21 pm
Volcano Plots showing fold-change and p-values of all genes



Fold-change (log2) versus significance (-log10 p-value) for each gene is shown. Significant genes (FDR < 0.05) are in blue.

October 17th 2018, 2:37:22 pm
Volcano Plots showing fold-change and p-values of all genes

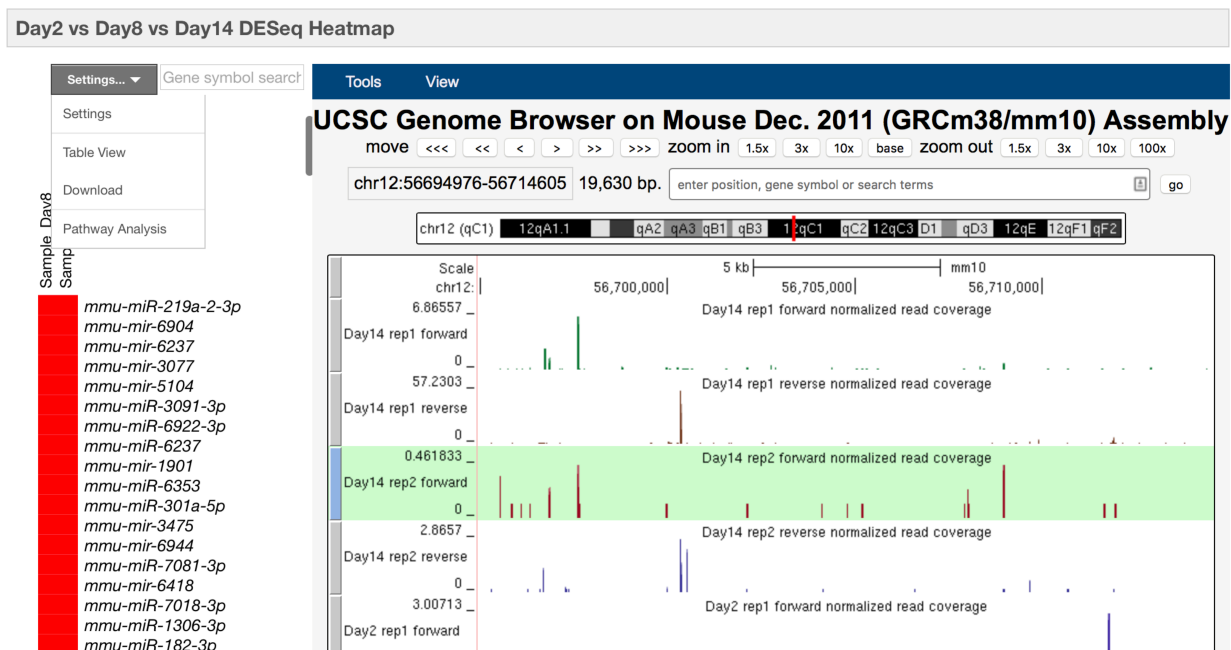


Fold-change (log2) versus significance (-log10 p-value) for each gene is shown. Significant genes (FDR < 0.05) are in blue.

October 17th 2018, 2:36:08 pm
Gene expression heatmap from DESeq output

[View Day2 vs Day8 vs Day14 DESeq Heatmap](#)

Interactive DESeq Heatmap



Users can select the desired gene feature on the interactive heatmap to explore its expression profile and read coverage on the UCSC Genome Browser. By hovering a gene feature on the heatmap, the expression data including log2 fold change, abundance, and adjusted P-value will be displayed on a popup window.

Gene symbol: [mmu-mir-6969](#) X
 Description: miRNA
 Locus: chrom 17 / 28,558,441 - 28,558,501

Sample Name	Log2 Fold Change	Abundance(smpl / ctrl)	P-Value
Sample_Day8/Sample_Day14	0.20	2.04 / 1.78	1.0
Sample_Day2/Sample_Day14	-Infinity	0.00 / 1.78	1.0

(row 123 of 342)

To view the read coverage of a desired gene feature on the genome browser, click the gene on the heatmap and the corresponding genomic region will be shown.

The heatmap data can be displayed in a table format by selecting Table View under the Settings menu. User can also download the expression data in display by selecting the Download option. To change the display and filtering options of the heatmap, choose the Settings menu item.

Settings ▼

▼ Sample

Control

Order

Sample_Day14 ▼

Apply

▸ Sorting

▼ Filters

Filter	Using	Condition	Value	
RNA type			<div style="border: 1px solid #ccc; padding: 2px; margin-bottom: 5px;"> RefSeq RefSeq_antisense miRNA other_ncRNA </div>	x
<div style="border: 1px solid #ccc; padding: 2px; border-radius: 5px; display: flex; align-items: center;"> Add filter ▼ </div>				

Apply

▸ Colors

▸ Gene Lists

▸ Genome Browser

▼ Gene Information

Show Gene Symbols:

On

Default Settings

By default, P-value < 0.05 is set as a filter to display data on the heatmap. If too few genes are shown, users can select to relax or remove the filter. Pairwise differential expression analysis is included. Users can select the appropriate sample as the control for comparison on the heatmap. Users can also upload a pre-determined gene list to be displayed on the heatmap.

For details on how to use the UCSC Genome Browser, please go to <https://genome.ucsc.edu/training/index.html>.