Galaxy Training!

Whole transcriptome analysis of Arabidopsis thaliana

Authors: Cristóbal Gallardo Pavankumar Videm Beatriz Serrano-Solano

Overview

Questions:

- Which miRNAs are upregulated in response to brassinosteroids?
- Which genes are potential target of brassinosteroid-induced miRNAs?

Objectives:

- Perform miRNA differential expression analysis
- Understand the quasi-mapping-based Salmon method for quantifying the expression of transcripts using RNA-Seq data
- Idenfity potential miRNAs involved in brassinosteroid-mediated regulation networks

Requirements:

- Introduction to Galaxy Analyses
- Sequence analysis
  - Quality Control: slides - hands-on
  - Mapping: slides - hands-on

Time estimation: 2 hours

Supporting Materials:

Slides Datasets Workflows FAQs Recordings

Available on these Galaxies

Last modification: Jan 19, 2023

Introduction

As sessile organisms, the survival of plants under adverse environmental conditions depends, to a large extent, on their ability to perceive stress stimuli and respond appropriately to counteract the potentially damaging effects. Coordination of phytohormones and reactive oxygen species are considered a key element for enhancing stress resistance, allowing fine-tuning of gene expression in response to environmental changes (Planas-Riverola *et al.* 2019, Ivashuta *et al.* 2011). These molecules constitute complex signalling networks, endowing with the ability to respond to a variable natural environment.

Brassinosteroids (BRs) are a group of plant steroid hormones essential for plant growth and development, as well as for controlling abiotic and biotic stress. Structurally, BRs are polyhydroxylated sterol derivatives with close similarity to animal hormones (Figure 1). This group of phytohormones is comprised of around 60 different compounds, of which brassinolide (BL), 24-epibrassinolide (EBR), and 28-homobrassinolide (HBR) are considered the most bioactive.

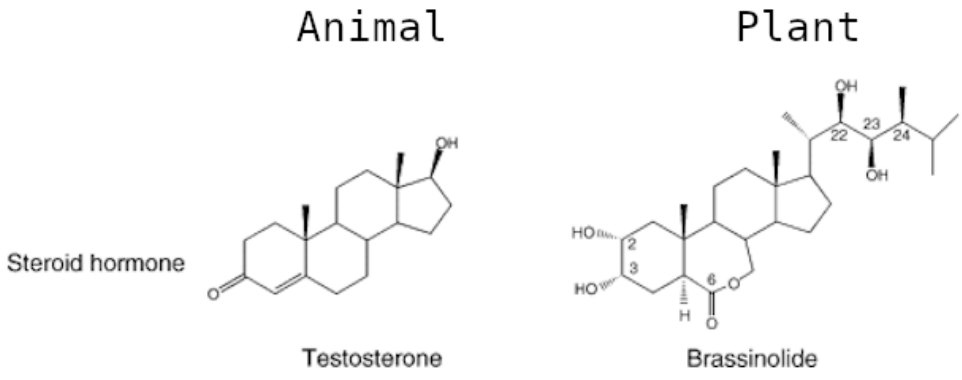


Figure 1: Structure of various plant and animal steroid hormones.

Several recent studies suggest that the BR-mediated gene regulatory networks have the potential to reshape the future of agriculture, not only by alleviating the antagonistic effect diverse abiotic stress conditios, such as drought, but also by enhancing plant growth and yield. For instance, in tomato (*Solanum lycopersicum*), EBR treatment enhances drought tolerance, improving photosynthetic capacity, leaf water status, and antioxidant defense (Wang *et al.* 2018). In pepper (*Capsicum annuum*), BL treatment increased the efficiency of light utilization under drought (Hu *et*

al. 2013). Gram (*Cicer arietinum*) plants exposed to drought stress and treated with BL showed significant increases in weight (Anwar *et al.* 2018). However, the mechanisms of BRs action in enhancing plant tolerance to abiotic stresses still remain largely unknown.

MicroRNAs (miRNAs), mainly 20–22 nucleotide small RNAs (sRNAs), are characterized for regulating gene expression at the post-transcriptional level. miRNAs are distinguished from other sRNAs by being generated from precursor harboring an imperfect stem-loop structure. Unlike in animals, the pre-processing of plant miRNA occurs in the nucleus (Figure 2). The pre-miRNAs are then exported to the cytoplasm after methylation and incorporated into the Argonaute 1 protein to form RISC (RNA-induced silencing complex). The miRNA itself does not have the ability to cleave mRNAs or interfere with translation, but it plays a role in scanning the appropriate target.

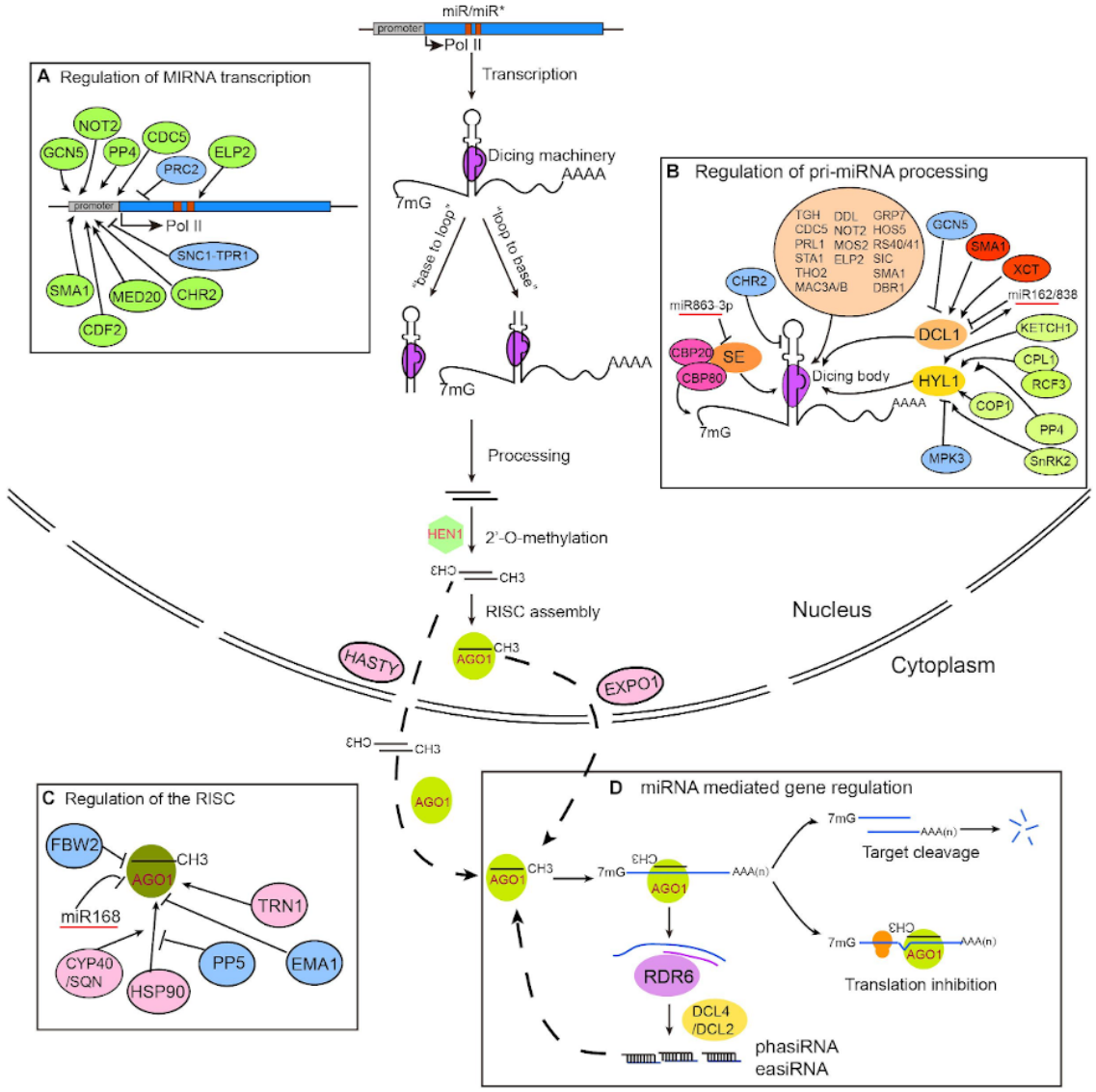


Figure 2: Plant miRNA biosynthesis, homeostasis and mechanisms of action (Wang *et al.* 2018).

miRNAs have been found to be important regulators of many physiological processes, such as stress and hormonal responses. Four factors justify the miRNAs to be considered as master regulators of the plant response to the surrounding environment:

- Multiple miRNA genes are regulated under given environmental conditions
- Computational predictions estimate that each miRNA regulates hundreds of genes

- The majority of plant miRNAs regulate genes encoding for transcription factors (TFs)
- Targets include not only mRNAs but also long noncoding RNAs (lncRNAs)

In plants, miRNAs can silence targets through RNA degradation as well as translational repression pathways, and unlike animals, a large proportion of miRNA and their targets have less than four mismatches. This feature has been exploited for developing miRNAs target prediction tools, providing an efficient approach to elucidate the miRNA-mediated regulatory networks, which can contribute to biotechnological solutions to improve crops productivity.

In this tutorial, inspired by Park *et al.* 2020, we aim to explore the interplay between brassinosteroids and the miRNA-gene silencing pathway, considered one of the most versatile regulatory mechanisms in response to stressful situations in plants.

Agenda

In this tutorial, we will cover:

Experimental design

The main objective of this training is to identify potential targets of miRNAs whose expression is induced by brassinosteroids. Our starting hypothesis is that there must be brassinosteroid-induced miRNAs that have high sequence complementarity with mRNAs whose expression is inhibited in the presence of these hormones (Figure 3).

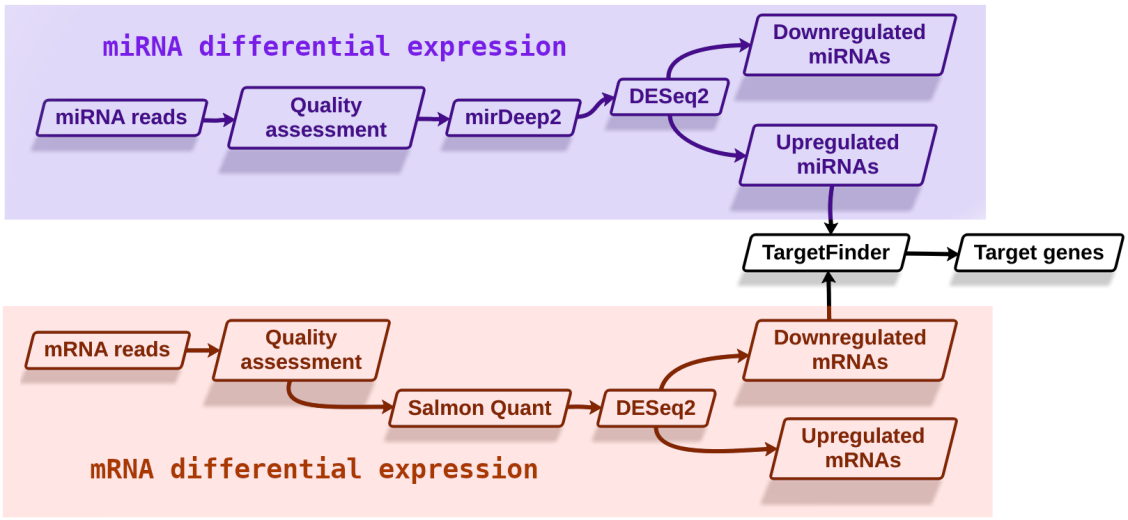


Figure 3: Experimental design

Background on data

The datasets to be used in this training can be classified into three groups: miRNA reads, mRNA reads and additional datasets.

miRNA reads

The miRNA datasets consist of six FASTQ files, obtained by using the Illumina GAxII sequencing platform. The plant samples were obtained from wild-type Ws-2 seedlings treated with mock or 1 µM EBR for 90 min before harvest. The original datasets are available in the NCBI SRA database, with the accession number SRP258575. As in the previous case, for this tutorial, we will use a reduced version of the data.

mRNA reads

The mRNA datasets consist of four FASTQ files, generated through the Illumina HiSeq 2000 sequencing system. The samples were obtained from wild-type Columbia (Col-0) seedlings treated with mock or 100 nM BL for 4 hours. The original datasets are available in the NCBI SRA database, with the accession number SRP032274. For this tutorial, subsets from the original data were generated in order to reduce the analysis run time.

Additional datasets

In addition to the RNA-Seq reads obtained from the NCBI database, we will use datasets from two sources:

- AtRTD2 is a high-quality transcript reference dataset developed to exploit the accuracy of transcript quantification tools such as **Salmon** and **Kallisto** in analyzing *Arabidopsis* RNA-Seq data.

- PmiREN is a comprehensive functional plant miRNA database that includes more than 20,000 annotated miRNAs diverse plant species.

Get data

The first step of our analysis consists of retrieving the miRNA-Seq datasets from Zenodo and organizing them into collections.

Hands-on: Retrieve miRNA-Seq and mRNA-Seq datasets

1. Create a new history for this tutorial
2. Import the files from Zenodo:
    - Open the file **upload** menu
    - Click on **Rule-based** tab
    - *"Upload data as"*: `Collection(s)`
    - Copy the following tabular data, paste it into the textbox and press `Build`

```
SRR11611349     Control miRNA
https://zenodo.org/record/4710649/files/SRR11611349_MIRNASEQ_CTL.fast
qsanger.gz      fastqsanger.gz
SRR11611350     Control miRNA
https://zenodo.org/record/4710649/files/SRR11611350_MIRNASEQ_CTL.fast
qsanger.gz      fastqsanger.gz
SRR11611351     Control miRNA
https://zenodo.org/record/4710649/files/SRR11611351.MIRNASEQ_CTLfastq
sanger.gz       fastqsanger.gz
SRR11611352     BR treated miRNA
https://zenodo.org/record/4710649/files/SRR11611352_MIRNASEQ_BL.fastq
sanger.gz       fastqsanger.gz
SRR11611353     BR treated miRNA
https://zenodo.org/record/4710649/files/SRR11611353_MIRNASEQ_BL.fastq
sanger.gz       fastqsanger.gz
SRR11611354     BR treated miRNA
https://zenodo.org/record/4710649/files/SRR11611354_MIRNASEQ_BL.fastq
sanger.gz       fastqsanger.gz
SRR1019436      Control mRNA
https://zenodo.org/record/4710649/files/SRR1019436_RNASEQ_CTL.fastqsa
nger.gz fastqsanger.gz
SRR1019437      Control mRNA
https://zenodo.org/record/4710649/files/SRR1019437_RNASEQ_CTL.fastqsa
nger.gz fastqsanger.gz
SRR1019438      BR treated mRNA
https://zenodo.org/record/4710649/files/SRR1019438_RNASEQ_BL.fastqsan
ger.gz  fastqsanger.gz
SRR1019439      BR treated mRNA
https://zenodo.org/record/4710649/files/SRR1019439_RNASEQ_BL.fastqsan
ger.gz  fastqsanger.gz
```

    - From **Rules** menu select `Add / Modify Column Definitions`
        - Click `Add Definition` button and select `List Identifier(s)`: column `A`

> Tip: Can't find *List Identifier*?

        - Click `Add Definition` button and select `Collection Name`: column `B`
        - Click `Add Definition` button and select `URL`: column `C`

- Click `Add Definition` button and select `Type`: column `D`
- Click `Apply` and press `Upload`

Tip: Adding a tag

Link to here | FAQs | Gitter Chat | Help Forum

Next we will retrieve the remaining datasets.

Hands-on: Retrieve the additional datasets

1. Import the files from Zenodo:
   - Open the file **upload** menu
   - *"Upload data as"*: `Datasets`
   - Once again, copy the tabular data, paste it into the textbox and press `Build`

```
annotation_AtRTD2.gtf
https://zenodo.org/record/4710649/files/annotation_AtRTD2_19April2016.gtf.gz
transcriptome.fasta
https://zenodo.org/record/4710649/files/transcriptome_AtRTD2_12April2016.fasta.gz
star_miRNA_seq.fasta
https://zenodo.org/record/4710649/files/star_miRNA_seq.fasta
mature_miRNA_AT.fasta
https://zenodo.org/record/4710649/files/mature_miRNA_AT.fasta
miRNA_stem-loop_seq.fasta
https://zenodo.org/record/4710649/files/miRNA_stem-loop_seq.fasta
```

   - From **Rules** menu select `Add / Modify Column Definitions`
     - Click `Add Definition` button and select `Name`: column `A`
     - Click `Add Definition` button and select `URL`: column `B`
   - Click `Apply` and press `Upload`

Link to here | FAQs | Gitter Chat | Help Forum

Details: Dataset subsampling

miRNA data analysis

Once we have imported the data, we can begin to study how brassinosteroid exposure alters gene expression patterns.

Quality assessment of miRNA reads

Due to technical limitations, sequencing is considered an error-prone process. In Illumina sequencing platforms, substitution type miscalls are the dominant source of errors, which can lead to inconsistent results. Another factor that can interfere with our analyses is the presence of adapter contaminants, which can result in an increased number of unaligned reads, since the adapter sequences are synthetic and do not occur in the genomic sequence.

Sequence quality control is therefore an essential first step in your analysis. We will use two popular tools for evaluating the quality of our raw reads: **FastQC** and **MultiQC**.

Comment

In order to visualize the data from both collections together in the **MultiQC** tool, it will be necessary to combine the results generated by **FastQC**. For more information on the topic of quality control, please see our dedicated training materials.

Hands-on: Initial quality check

1. **FastQC** with the following parameters:
   - *"Dataset collection"*: `Control miRNA`
2. Rename the outputs as `FastQC unprocessed control miRNA: RawData` and `FastQC unprocessed control miRNA: Webpage`
3. Repeat the previous steps with the `BR treated miRNA` collection.
4. **Merge Collections** with the following parameters:
   - In *"Input collections"*:
     - *"1.Input Collections"*: `FastQC unprocessed control miRNA: RawData`
     - *"2.Input Collections"*: `FastQC unprocessed BR treated mRiNA: RawData`
5. **MultiQC** with the following parameters:
   - In *"Results"*:
     - *"Which tool was used generate logs?"*: `FastQC`
     - *"Dataset collection"*: select the output generated in the previous step.
   - In *"Report title"*: `miRNA initial quality check`
6. Click on the (eye) icon and inspect the generated HTML file

Link to here | FAQs | Gitter Chat | Help Forum

Question

Based on the information provided by **MultiQC**, is it necessary to trimming/filtering the reads?

Solution

To remove the adapters contamination, we will employ the **Trim Galore** tool, a wrapper script around **Cutadapt** and **FastQC** which automates quality and adapter trimming.

Hands-on: Trimming of adapter sequences

1. **Trim Galore!** with the following parameters:
   - *"Is this library paired- or single-end?"*: `Single-end`
     - *"Reads in FASTQ format"*: `Control miRNA`
     - *"Adapter sequence to be trimmed"*: `Illumina small RNA adapters`
2. Rename the output collection as `Control miRNA trimmed`
3. Repeat the previous steps with the `BR treated miRNA` collection.

Link to here | FAQs | Gitter Chat | Help Forum

Next, we will reassess the quality of the sequences to check if the adapters have been removed.

Hands-on: Post-processing quality check

1. **FastQC** with the following parameters:
   - *"Dataset collection"*: `Control miRNA trimmed`
2. Rename the outputs as `FastQC processed control miRNA: RawData` and `FastQC processed control miRNA: Webpage`
3. Repeat the previous steps with the `BR treated miRNA trimmed` collection.
4. **Merge Collections** with the following parameters:
   - In *"Input collections"*:
     - *"1.Input Collections"*: `FastQC processed control miRNA: RawData`
     - *"2.Input Collections"*: `FastQC processed BR treated miRNA: RawData`
5. **MultiQC** with the following parameters:
   - In *"Results"*:
     - *"Which tool was used generate logs?"*: `FastQC`
     - *"Dataset collection"*: select the output generated in the previous step.
   - In *"Report title"*: `miRNA trimmed quality check`
6. Click on the (eye) icon and inspect the generated HTML file

Link to here | FAQs | Gitter Chat | Help Forum

The evaluation of the report generated by **MultiQC** after having processed the samples by **Trim Galore** indicates that the G/C content has been successfully corrected, and that the adapter contamination has been eliminated. However, the samples still show a high degree of over-represented sequences (Figure 7).

Figure 7: Report of overexpressed sequences in the miRNA reads

Question

What do you think could be the cause of the high number of over-represented sequences?

Solution

miRNA quantification: **MiRDeep2**

Quantification of miRNAs requires to use two different tools:

- The **MiRDeep2 Mapper** tool for preprocessing the reads.
- The **MiRDeep2 Quantifier** tool for mapping the deep sequencing reads to predefined miRNA precursors and determining the expression of the corresponding miRNAs. It is carried out in two steps: firstly, the predefined mature miRNA sequences are mapped to the predefined precursors (optionally, predefined star sequences can be mapped to the precursors too). And second, the deep sequencing reads are mapped to the precursors.

Hands-on: Quantification of miRNAs

1. **MiRDeep2 Mapper** with the following parameters:
   - *"Deep sequencing reads"*: `Control miRNA trimmed`
   - *"Remove reads with non-standard nucleotides"*: `Yes`
   - *"Clip 3' Adapter Sequence"*: `Don't Clip`
   - *"Collapse reads and/or Map"*: `Collapse`
2. Rename the output as `Collapsed control miRNA`
3. Repeat the previous stages by providing `BR treated miRNA trimmed` as input, and rename it as `Collapsed BR treated miRNA`.
4. **MiRDeep2 Quantifier** with the following parameters:
   - *"Collapsed deep sequencing reads"*: `Collapsed control miRNA`
   - *"Precursor sequences"*: `miRNA_stem-loop_seq.fasta`
   - *"Mature miRNA sequences"*: `mature_miRNA_AT.fasta`
   - *"Star sequences"*: `star_miRNA_seq.fasta`
5. Rename the outputs as `MiRDeep2 control miRNA` and `MiRDeep2 control miRNA (html)`.
6. Repeat the fourth step by providing `Collapsed BR treated miRNA` as input, and rename the outputs as `MiRDeep2 BR treated miRNA` and `MiRDeep2 BR treated miRNA (html)`

Link to here | FAQs | Gitter Chat | Help Forum

To use the outputs generated by **MiRDeep2 Quantifier** in the differential expression analysis, it is necessary to modify the datasets.

Hands-on: Edition of MiRDeep2 Quantifier outputs

1. **Cut columns from a table** with the following parameters:
   - *"Cut columns"*: `c1,c2`
   - *"Delimited by"*: `Tab`
   - *"From"*: `MiRDeep2 control miRNA`
2. Rename the output as `control miRNA counts`
3. **Cut columns from a table** with the following parameters:
   - *"Cut columns"*: `c1,c2`
   - *"Delimited by"*: `Tab`
   - *"From"*: `MiRDeep2 BR treated miRNA`
4. Rename the output as `BR treated miRNA counts`

Link to here | FAQs | Gitter Chat | Help Forum

Differential expression analysis of miRNAs: **DESeq2**

DESeq2 is a tool for differential gene expression analysis based on a negative binomial generalized linear model. **DESeq2** internally corrects the differences in library size, due to which no preliminary normalization of input datasets is required.

> 💬 Comment
>
> It is desirable to use at least three replicates of each experimental condition to ensure sufficient statistical power.

> 🖉 Hands-on: Differential expression analysis using DESeq2
>
> 1. **DESeq2** with the following parameters:
>   - *"How"*: `Select datasets per level`
>   - In *"Factor"*:
>   - ⊞ *"Insert Factor"*
>   - *"Specify a factor name, e.g. effects_drug_x or cancer_markers"*: `effects_brassinolide`
>   - In *"Factor level"*:
>   - ⊞ *"Insert Factor level"*
>   - *"Specify a factor level"*: `brassinolide`
>   - 🗀 *"Counts file(s)"*: `BR treated miRNA counts`
>   - ⊞ *"Insert Factor level"*
>   - *"Specify a factor level"*: `control`
>   - 🗀 *"Counts file(s)"*: `control miRNA counts`
>   - *"Files have header?"*: `No`
>   - *"Choice of Input data"*: `Count data (e.g. from HTSeq-count, featureCounts or StringTie)`
> 2. Rename the outputs as `DESeq2 results miRNA` and `DESeq2 plots miRNA`
> 3. Click on the 👁 (eye) icon and inspect the `DESeq2 plots miRNA` file
>
> Link to here | ? FAQs | Gitter Chat | Help Forum

DESeq2 generated 2 outputs: a table with the normalized counts and a graphical summary of the results. To evaluate the similarity of our samples, we are going to inspect the Principal Component Analysis (PCA) plot. PCA allows evaluating the dominant directions of the highest variability in the data. Thus, the samples subjected to the same conditions should cluster together.

Figure 8: PCA plot of expression data from control and BR treated miRNA samples.

As can be seen, the main axes account for only 47% and 19% of the total variation. This suggests that the effect of brassinosteroids on miRNA regulation is limited (Figure 8).

Filter significantly differentially expressed miRNAs

Next, we will extract those genes whose expression is statistically significantly differentially expressed (DE) due to BR treatment by selecting those whose adjusted p-value is less than or 0.05. A cut-off value of 0.05 indicates that the probability of a false positive result is less than 5%.

Details: p-value vs p-adj

Hands-on: Filter differentially expressed miRNAs

1. **Filter** data on any column using simple expressions (Galaxy Version 1.1.1) with the following parameters:
   - *"Filter"*: `DESeq2 results miRNA`
   - *"With following condition"*: `c7<0.05`

Link to here | FAQs | Gitter Chat | Help Forum

Question

How many miRNAs show statistically significant differential expression?

Solution

To get any sensible results, it is worth analyzing the full dataset. You can analyze the full datasets following the above tutorial. Otherwise, you can import the DESeq2 analysis results that we generated from full dataset into your history.

Hands-on: Retrieve the DESeq2 analysis results on full miRNA dataset

1. Import the files from Zenodo:
   - Open the file **upload** menu
   - Click of the **Paste/Fetch** button
   - Copy the Zenodo links and press `Start`

     ```
     https://zenodo.org/record/4663389/files/miRNA_DESeq2_results_complete_dataset.tabular
     ```

2. Rename each dataset according to the sample id (e.g. `miRNA_DESeq2_results_complete_dataset`)
3. Add all miRNA data analysis related tags: #miRNA #BR #control

Link to here | FAQs | Gitter Chat | Help Forum

Repeat the filtering step on the imported DESeq2 result.

Hands-on: Filter and sort differentially expressed miRNAs from the full dataset

1. **Filter** data on any column using simple expressions (Galaxy Version 1.1.1) with the following parameters:
   - *"Filter"*: `miRNA_DESeq2_results_complete_dataset`
   - *"With following condition"*: `c7<0.05`
2. Rename the output as `Differentially expressed miRNAs`
3. **Filter** data on any column using simple expressions (Galaxy Version 1.1.1) with the following parameters:
   - *"Filter"*: `Differentially expressed miRNAs`
   - *"With following condition"*: `c3>0`
4. Rename the output as `Upregulated miRNAs`
5. **Sort** data in ascending or descending order (Galaxy Version 1.1.0) with the following parameters:
   - *"Sort Dataset"*: `Upregulated miRNAs`
   - *"on column"*: `Column: 3`
   - *"everything in"*: `Descending order`

Link to here | FAQs | Gitter Chat | Help Forum

Question

1. How many miRNAs are differentially expressed?
2. How many miRNAs show statistically significant upregulation and what is the most upregulated miRNA?

Solution

mRNA data analysis

Once the differential expression analysis of miRNAs has been carried out, the next stage is to analyze how mRNA expression is altered in response to brassinosteroids.

Quality assessment of mRNA reads

As in the previous section, we shall begin by assessing the quality of our sequences.

Hands-on: Initial quality check

1. **FastQC** with the following parameters:
   - *"Dataset collection"*: `Control mRNA`
2. Rename the outputs as `FastQC unprocessed control mRNA: RawData` and `FastQC unprocessed control mRNA: Webpage`
3. Repeat the previous step with the `BR treated mRNA` collection.
4. **Merge Collections** with the following parameters:
   - In *"Input collections"*:
     - *"1.Input Collections"*: `FastQC unprocessed control mRNA: RawData`
     - *"2.Input Collections"*: `FastQC unprocessed BR treated mRNA: RawData`
5. **MultiQC** with the following parameters:
   - In *"Results"*:
     - *"Which tool was used generate logs?"*: `FastQC`
     - *"Dataset collection"*: select the output generated in the previous step.
   - *"Report title"*: `mRNA initial quality check`
6. Click on the (eye) icon and inspect the generated HTML file

Link to here | FAQs | Gitter Chat | Help Forum

Question

Are there any stats that indicate the need to process the samples in order to improve their quality?

Solution

Comment

Although our samples have adapters, we are not going to trim them in this case. We will explain the reason in the next section.

Quantification of gene expression: **Salmon**

After performing the quality assessment of the reads, we can move on to quantifying the gene expression. The aim of this step is to identify which transcript each read comes from and the total number of reads associated with each transcript. In this tutorial, we will use the Salmon tool (Patro *et al.* 2017) for the quantification of mRNA transcripts.

One of the characteristics of **Salmon** is that it doesn't require performing a base-to-base alignment, which is the time-consuming step of tools such as **STAR** and **HISAT2**. Salmon relies on the quasi-mapping concept, a new mapping technique that allows the rapid and accurate mapping of RNA-Seq reads to a target transcriptome. Rather than a standard alignment, quasi-mapping seeks to find the best mappings for each read, and does so by finding minimal collections of dynamically sized, right-maximal, matching contexts between target and query positions (Srivastava *et al.* 2016)

The quasi-mapping approach by **Salmon** requires a reference index to determine the position and orientation information for accurate mapping prior to quantification. It allows providing the transcriptome in a format that optimizes its use in transcript identification and quantification.

After determining the best mapping for each read, **Salmon** generates the final transcript abundance estimation after modeling sample-specific parameters and biases. Reads that map equally well to more than one transcript will have the count divided between all of the mappings, thus avoiding the loss of information on the different gene isoforms.

Details: Quasi-mapping algorithm

Hands-on: Quantify gene expression with Salmon

1. **Salmon quant** with the following parameters:
    - *"Select salmon quantification mode:"*: `Reads`
        - *"Select a reference transcriptome from your history or use a built-in index?"*: `Use one from the history`
            - In *"Salmon index"*:
                - *"Transcripts fasta file"*: `transcriptome.fasta`
        - In *"Data input"*:
            - *"Is this library mate-paired?"*: `Single-end`
                - *"FASTQ/FASTA file"*: `Control mRNA`
        - *"Validate mappings"*: `Yes`
    - *"File containing a mapping of transcripts to genes"*: `annotation_AtRTD2.gtf`
2. Rename the outputs as `Salmon control mRNA (Quantification)` and `Salmon control mRNA (Gene Quantification)`
3. Repeat the previous procedure by using the `BR treated mRNA` dataset

Comment: Quasi-mapping sequence requirements

Trimming the reads is not required when using this method, since if there are k-mers in the reads that are not in the hash table, they are not counted. Quantification of the reads is only as good as the quality of the reference transcriptome.

Link to here | FAQs | Gitter Chat | Help Forum

Salmon generates two outputs for each input sample:

- Quantification: summarizes the quantifications by transcript
- Gene quantification: summarizes the quantification by gene

Each output consists of a tabular dataset with five columns:

Field	Description
Name	The name of the target transcript provided in the input transcriptome.
Length	The length of the target transcript.
EffectiveLength	The computed effective length.
TPM	The relative abundance of this transcript in units of Transcripts Per Million.
NumReads	The number of reads mapping to each transcript that was quantified.

Question

Could you explain why we did not trim the reads before?

Solution

Differential expression analysis of mRNAs: **DESeq2**

Now, let's analyze which genes show statistically significant differential expression in response to brassinosteoids.

Hands-on: Differential expression analysis using DEseq2

1. **DESeq2** with the following parameters:
    - *"How"*: `Select datasets per level`
        - In *"Factor"*:
            - *"Insert Factor"*
                - *"Specify a factor name, e.g. effects_drug_x or cancer_markers"*: `effects_brassinolide`
                - In *"Factor level"*:
                    - *"Insert Factor level"*
                        - *"Specify a factor level"*: `brassinolide`
                        - *"Counts file(s)"*: `Salmon BR treated mRNA (Gene Quantification)`
                    - *"Insert Factor level"*
                        - *"Specify a factor level"*: `control`
                        - *"Counts file(s)"*: `Salmon control mRNA (Gene Quantification)`
    - *"Choice of Input data"*: `TPM values (e.g. from kallisto, sailfish or salmon)`
        - *"Program used to generate TPMs"*: `Salmon`
        - *"Gene mapping format"*: `GTF/GFF3`
            - *"GTF/GFF3 annotation file"*: `annotation_AtRTD2.gtf`
2. Rename the outputs as `DESeq2 results mRNA` and `DESeq2 plots mRNA`

Link to here | FAQs | Gitter Chat | Help Forum

Figure 11: PCA plot of differential expression data from control and BR treated samples.

Question

What conclusions about the similarity of the samples can be derived from the PCA plot?

Solution

Filter significantly differentially expressed mRNAs

To conclude the analysis of the differential expression of mRNAs, we will extract those transcripts that show a significant induction of expression in response to brassinosteroids. Before continuing with further analysis, similar to miRNA data analysis, import the DESeq2 results generated from full mRNA datasets.

Hands-on: Retrieve the DESeq2 analysis results on full mRNA dataset

1. Import the files from Zenodo:
   - Open the file **upload** menu
   - Click of the **Paste/Fetch** button
   - Copy the Zenodo links and press `Start`

     ```
     https://zenodo.org/record/4663389/files/mRNA_DESeq2_results_complete_dataset.tabular
     ```

2. Rename each dataset according to the sample id (e.g. `mRNA_DESeq2_results_complete_dataset`)
3. Add all mRNA data analysis related tags: #mRNA #BR #control

Link to here | FAQs | Gitter Chat | Help Forum

Now we continue with the DE genes analysis.

Hands-on: Extract significantly DE genes

1. **Filter** data on any column using simple expressions (Galaxy Version 1.1.1) with the following parameters:
   - *"Filter"*: `mRNA_DESeq2_results_complete_dataset`
   - *"With following condition"*: `c7<0.05`
2. Rename the output as `Differentially expressed mRNAs`
3. **Filter** data on any column using simple expressions (Galaxy Version 1.1.1) with the following parameters:
   - *"Filter"*: `Differentially expressed mRNAs`
   - *"With following condition"*: `c3>1`
4. Rename it as `Upregulated mRNAs`
5. **Filter** data on any column using simple expressions (Galaxy Version 1.1.1) with the following parameters:
   - *"Filter"*: `Differentially expressed mRNAs`
   - *"With following condition"*: `c3<-1`
6. Rename it as `Downregulated mRNAs`

Link to here | FAQs | Gitter Chat | Help Forum

Question

1. How many genes show statistically significant differential expression?
2. How many of them are significantly upregulated (at least two-fold change)? And downregulated?
3. What is the most significantly DE downregulated gene and what is it biological function?

Solution

Identification of miRNA targets

To predict which miRNAs target which mRNAs, first we need their transcriptomic sequences in FASTA format. Now we will obtain the sequences of miRNAs whose expression is induced by brassinosteroids.

Hands-on: Obtaining the sequences of upregulated miRNAs

> Comment
>
> The following tools can be found in *Text Manipulation* and *Filter and Sort* sections.

1. **Cut columns from a table** with the following parameters:
   - *"Cut columns"*: `c1`
   - *"Delimited by"*: `Tab`
   - *"From"*: `Upregulated miRNAs`
2. Rename the output as `Upregulated miRNA ids`
3. **Filter FASTA** with the following parameters:
   - *"FASTA sequences"*: `mature_miRNA_AT.fasta`
   - *"Criteria for filtering on the headers"*: `List of IDs`
     - *"List of IDs to extract sequences for"*: `Upregulated miRNA ids`
     - *"Match IDs by"*: `Default: ID is expected at the beginning: >ID`
4. **Filter FASTA** with the following parameters:
   - *"FASTA sequences"*: `star_miRNA_seq.fasta`
   - *"Criteria for filtering on the headers"*: `List of IDs`
     - *"List of IDs to extract sequences for"*: `Upregulated miRNA ids`
     - *"Match IDs by"*: `Default: ID is expected at the beginning: >ID`
5. **Concatenate datasets** tail-to-head (Galaxy Version 1.0.0) with the following parameters:
   - *"Concatenate Dataset"*: output of **Filter FASTA** on `mature_miRNA_AT.fasta`
   - Insert Dataset
     - *"Select"*: output of **Filter FASTA** on `star_miRNA_seq.fasta`
6. Rename it as `Upregulated miRNA sequences`
7. Click on the (eye) icon and inspect the `Upregulated miRNA sequences` file

Link to here | FAQs | Gitter Chat | Help Forum

To identify putative targets of upregulated miRNAs, it is necessary to obtain the sequences of all downregulated mRNAs in FASTA format.

Hands-on: Obtaining the gene sequences of downregulated mRNAs

1. **Cut columns from a table** with the following parameters:
   - *"Cut columns"*: `c1`
   - *"Delimited by"*: `Tab`
   - *"From"*: `Downregulated mRNAs`
2. Rename the output as `Downregulated mRNA ids`
3. **Filter FASTA** with the following parameters:
   - *"FASTA sequences"*: `transcriptome.fasta` (Input dataset)
   - *"Criteria for filtering on the headers"*: `List of IDs`
     - *"List of IDs to extract sequences for: `Downregulated mRNA ids`
     - *"Match IDs by"*: `Custom regex pattern`
       - *"Regex search pattern for ID"*: `GENE=(AT.{7})`
4. Rename it as `Downregulated mRNA sequences`
5. Click on the (eye) icon and inspect the `Downregulated mRNA sequences` file

Link to here | FAQs | Gitter Chat | Help Forum

miRNA target prediction using **TargetFinder**

We are now ready to launch the search for miRNA target genes. For this we will use the **TargetFinder** tool that is implemented and used for miRNA target prediction in plants (Srivastava *et al.* 2014, Ma *et al.* 2017).

Hands-on: identification of potential miRNA targets

1. **TargetFinder** with the following parameters:
   - *"Input small RNA sequences file"*: `Upreguled miRNA sequences`
   - *"Target sequence database file"*: `Downregulated mRNA sequences`
   - *"Prediction score cutoff value"*: `5.0`
   - *"Output format"*: `Tab-delimited format`
2. Click on the (eye) icon and inspect the output of **TargetFinder**.

Link to here | FAQs | Gitter Chat | Help Forum

Congratulations! We have identified the following 5 potential genes involved in the brassinosteroid-miRNA regulatory network.

Figure 12: TargetFinder results.

Finally, we can access all the information available on the genes identified in the TAIR database:

- AT5G10180: ARABIDOPSIS SULFATE TRANSPORTER 68, AST6
- AT3G09220: LAC7, LACCASE 7
- AT2G46850: Protein kinase superfamily protein
- AT5G64260: EXL2, EXORDIUM LIKE 2

- AT3G63200: PATATIN-LIKE PROTEIN 9, PLA IIIB

Both AT4G14365 and AT1G26890 are not well characterized genes. In the case of AT5G50570, experimental research have demonstrated that this gene is involved in flooding tolerance in *Medicago sativa*, through a signaling path mediated by miR156 (Feyissa *et al.* 2021). On the other hand, according to Gao *et al.* 2017, SPL13 is likely a negative regulator of plant vegetative growth. The interaction between miR156 and SPL transcription factors has been suggested for diverse Poaceae family members (Yue *et al.* 2021).

💬 Comment: Hypothesis testing

One of the hypotheses that we could propose from our results is: the inhibition of the AT2G46850 gene can result in plants with improved resistance to drought conditions. Is it possible to validate it? Yes! We propose this approach: to acquire AT2G46850 mutant seeds and wild type seeds, grow them under two controlled conditions: watered and drought stress, and analyze plant weight after 33 days (Figure 13).

Figure 13: Arabidopsis growth conditions protocol (de Ollas *et al.* 2019).

Optional exercise

As additional activity, you can try to repeat the workflow by using the sequences stored in the NCBI GEO database with the accession number `GSE119382`. In that case, we will compare gene expression patterns of mutants overexpressing the brassinosteroid receptor BRL3 under two experimental conditions: control and drought-stress. The required datasets are available in the data library:

🛠 Hands-on: Import data from the Data Libraries

1. Go into **Shared data** (top panel) and click on **Data Libraries**
2. In the search box enter the following identifier: `4710649`
3. Select the following files:

```
https://zenodo.org/record/4710649/files/SRR7779222_BRL3_mRNA_drought.fastqsanger.gz
https://zenodo.org/record/4710649/files/SRR7779223_BRL3_mRNA_drought.fastqsanger.gz
https://zenodo.org/record/4710649/files/SRR7779224_BRL3_mRNA_drought.fastqsanger.gz
```

4. Click on **Export to History** and **as a Collection**
5. In the pop-up window press `Continue`
6. Provide it the name `BRL3 mRNA drought` and push `Create list`
7. Repeat the previous procedure with the remaining files:

```
https://zenodo.org/record/4710649/files/SRR7779228_BRL3_mRNA_watered.fastqsanger.gz
https://zenodo.org/record/4710649/files/SRR7779229_BRL3_mRNA_watered.fastqsanger.gz
```

8. Finally provide it the name `BRL3 mRNA control` and push `Create list`

We will use the upregulated miRNAs obtained in the previous analysis in order to identify potential targets.

```
Upregulated miRNA
https://zenodo.org/record/4710649/files/upregulated_miRNA_BR_complete_dataset.fasta
```

Conclusion

In this tutorial, we have analyzed RNA sequencing data to extract information about potential genes regulated by brassinosteroids. For this purpose, the approach chosen was the identification of genes complementary to miRNAs upregulated in response by brassinosteroids. The final result has been the identification of five potential miRNA targets.

Key points

- MiRDeep2 in combination with the PmiREN database allows to quantify miRNA expression in plants
- The use of the Salmon tool and the AtRTD2 transcriptome allows rapid and accurate quantification of transcripts in *Arabidopsis*
- The combination of differential expression analysis and the TargetFinder tool is an effective strategy for the identification of miRNA target genes

Frequently Asked Questions

Have questions about this tutorial? Check out the tutorial FAQ page or the FAQ page for the Transcriptomics topic to see if your question is listed there. If not, please ask your question on the GTN Gitter Channel or the Galaxy Help Forum

Useful literature

Further information, including links to documentation and original publications, regarding the tools, analysis techniques and the interpretation of results described in this tutorial can be found here.

References

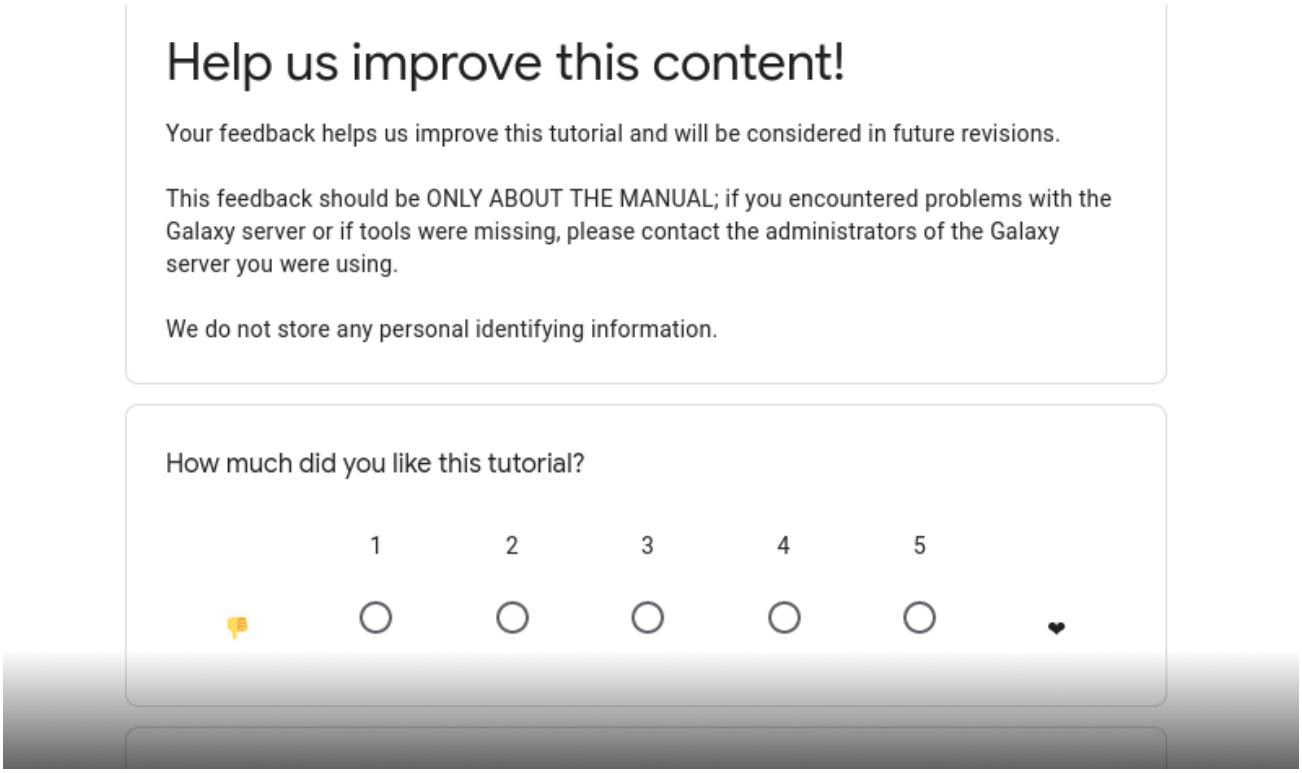
1. Glazov, E. A., S. McWilliam, W. C. Barris, and B. P. Dalrymple, 2008 **Origin, Evolution, and Biological Role of miRNA Cluster in DLK-DIO3 Genomic Region in Placental Mammals**. Molecular Biology and Evolution 25: 939–948. 10.1093/molbev/msn045
2. Ivashuta, S., I. R. Banks, B. E. Wiggins, Y. Zhang, T. E. Ziegler *et al.*, 2011 **Regulation of Gene Expression in Plants through miRNA Inactivation** (G. Dalmasso, Ed.). PLoS ONE 6: e21330. 10.1371/journal.pone.0021330
3. Hu, W.-hai, X.-hong Yan, Y.-an Xiao, J.-jun Zeng, H.-juan Qi *et al.*, 2013 **24-Epibrassinosteroid alleviate drought-induced inhibition of photosynthesis in Capsicum annuum**. Scientia Horticulturae 150: 232–237. 10.1016/j.scienta.2012.11.012
4. Srivastava, P. K., T. Moturu, P. Pandey, I. T. Baldwin, and S. P. Pandey, 2014 **A comparison of performance of plant miRNA target prediction tools and the characterization of features for genome-wide target prediction**. BMC Genomics 15: 348. 10.1186/1471-2164-15-348
5. Srivastava, A., H. Sarkar, N. Gupta, and R. Patro, 2016 **RapMap: a rapid, sensitive and accurate tool for mapping RNA-seq reads to transcriptomes**. Bioinformatics 32: i192–i200. 10.1093/bioinformatics/btw277
6. Gao, R., M. Y. Gruber, L. Amyot, and A. Hannoufa, 2017 **SPL13 regulates shoot branching and flowering time in Medicago sativa**. Plant Molecular Biology 96: 119–133. 10.1007/s11103-017-0683-8
7. Ma, X., C. Liu, L. Gu, B. Mo, X. Cao *et al.*, 2017 **TarHunter, a tool for predicting conserved microRNA targets and target mimics in plants** (I. Hofacker, Ed.). Bioinformatics 34: 1574–

1576. 10.1093/bioinformatics/btx797

8. Patro, R., G. Duggal, M. I. Love, R. A. Irizarry, and C. Kingsford, 2017 **Salmon provides fast and bias-aware quantification of transcript expression**. Nature Methods 14: 417–419. 10.1038/nmeth.4197
9. Anwar, A., Y. Liu, R. Dong, L. Bai, X. Yu *et al.*, 2018 **The physiological and molecular mechanism of brassinosteroid in response to stress: a review**. Biological Research 51: 10.1186/s40659-018-0195-2
10. Wang, C.-Y., T. Jégu, H.-P. Chu, H. J. Oh, and J. T. Lee, 2018 **SMCHD1 Merges Chromosome Compartments and Assists Formation of Super-Structures on the Inactive X**. Cell 174: 406–421.e25. 10.1016/j.cell.2018.05.007
11. Wang, Y., J.-J. Cao, K.-X. Wang, X.-J. Xia, K. Shi *et al.*, 2018 **BZR1 Mediates Brassinosteroid-Induced Autophagy and Nitrogen Starvation in Tomato**. Plant Physiology 179: 671–685. 10.1104/pp.18.01028
12. Planas-Riverola, A., A. Gupta, I. Betegón-Putze, N. Bosch, M. Ibañes *et al.*, 2019 **Brassinosteroid signaling in plant development and adaptation to stress**. Development 146: dev151894. 10.1242/dev.151894
13. Ollas, C. de, C. Segarra-Medina, M. González-Guzmán, J. Puertolas, and A. Gómez-Cadenas, 2019 **A customizable method to characterize Arabidopsis thaliana transpiration under drought conditions**. Plant Methods 15: 10.1186/s13007-019-0474-0
14. Park, S. Y., J.-H. Choi, D.-H. Oh, J. C. Johnson, M. Dassanayake *et al.*, 2020 **Genome-wide analysis of brassinosteroid responsive small RNAs in Arabidopsis thaliana**. Genes & Genomics 42: 957–969. 10.1007/s13258-020-00964-2
15. Feyissa, B. A., L. Amyot, V. Nasrollahi, Y. Papadopoulos, S. E. Kohalmi *et al.*, 2021 **Involvement of the miR156/SPL module in flooding response in Medicago sativa**. Scientific Reports 11: 10.1038/s41598-021-82450-7
16. Yue, E., H. Tao, and J. Xu, 2021 **Genome-wide analysis of microRNA156 and its targets, the genes encoding SQUAMOSA promoter-binding protein-like (SPL) transcription factors, in the grass family Poaceae**. Journal of Zhejiang University-SCIENCE B 22: 366–382. 10.1631/jzus.b2000519

Feedback

Did you use this material as an instructor? Feel free to give us feedback on how it went.
Did you use this material as a learner or student? Click the form below to leave feedback.



Citing this Tutorial

1. Cristóbal Gallardo, Pavankumar Videm, Beatriz Serrano-Solano, **Whole transcriptome analysis of Arabidopsis thaliana (Galaxy Training Materials)**. https://training.galaxyproject.org/training-material/topics/transcriptomics/tutorials/mirna-target-finder/tutorial.html Online; accessed TODAY
2. Batut et al., 2018 **Community-Driven Data Analysis Training for Biology** Cell Systems 10.1016/j.cels.2018.05.012

BibTeX

Congratulations on successfully completing this tutorial!